

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2004, 08:23:26 ; Search time 9384 Seconds

(without alignments)  
11225.734 Million cell updates/sec

Title: US-09-836-613-1

Perfect score: 2575

Sequence: 1 CCCGGGCTTACCTTCGGGT.....AGTCAGCGCGCCGGAATTC 2575

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 288871 segs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_ba:\*

2: gb\_hcg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_bcs:\*

12: gb\_ey:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_scs:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_hcg\_hum:\*

31: em\_hcg\_inv:\*

32: em\_hcg\_other:\*

33: em\_hcg\_mus:\*

34: em\_hcg\_pln:\*

35: em\_hcg\_rpd:\*

36: em\_hcg\_mam:\*

37: em\_hcg\_vrt:\*

38: em\_ey:\*

39: em\_hugo\_hum:\*

40: em\_hugo\_mus:\*

41: em\_hugo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	2575	100.0	2575	6	AR160672	AR160672 Sequence
2	2575	100.0	2575	9	US040846	U40846 Human alpha
3	2455	95.3	2458	9	HSU43573	U43573 Human alpha
4	2446	95.0	2446	9	HDMUFRSA	L78464 Human alpha
5	2063.6	80.1	2933	9	HDMUFRSA	L78465 Human alpha
6	1623.6	63.1	2452	10	MM085247	U85247 Mus musculus
7	1426.6	55.4	10127	9	HSU43572	U43572 Human alpha
8	1426.6	55.4	10380	6	AR160673	AR160673 Sequence
9	1426.6	55.4	46610	9	HSU34879	U34879 Human 17-be
10	1425	55.3	179146	9	AC067852	AC067852 Homo sapi
11	1415.6	55.0	114531	2	AC002537	AC002537 Homo sapi
12	1414.6	54.9	21764	9	HDM17BHD	M84472 Human 17-be
13	1114.2	43.3	2949	5	AF331668	AF331668 Dromaius
14	874.8	34.0	7721	10	AF003255	AF003255 Mus muscu
15	874.8	34.0	10938	10	AF363242	AF363242 Mus muscu
16	874.8	34.0	181557	10	AC069014	AC069014 Mus muscu
17	874.8	34.0	218666	2	EX323890	EX323890 Mus muscu
18	873.2	33.9	73844	10	EX255926	EX255926 Mouse DNA
19	865.2	33.6	235178	2	AC098921	AC098921 Rattus no
20	687.8	26.7	5762	10	RN17BHD	X98038 R. norvegicu
21	584.2	22.7	4217	5	AF331669	AF331669 Dromaius
22	369.4	14.3	2759	3	AY058738	AY058738 Drosophila
23	369.4	14.3	36890	2	AC018157	AC018157 Drosophila
24	369.4	14.3	79076	3	AC004423	AC004423 Drosophila
25	369.4	14.3	186490	3	AC092224	AC092224 Drosophila
26	369.4	14.3	308031	3	AE003621	AE003621 Drosophila
27	286.8	11.1	2452	8	AY117179	AY117179 Arabidops
28	286.8	11.1	2702	8	NTA18209	Y18209 Nicotiana t
29	258.2	10.0	2631	8	AY080811	AY080811 Arabidops
30	209.6	8.1	10545	1	AE005727	AE005727 Caulobact
31	183.4	7.1	298550	1	AP005029	AP005029 Streptomy
32	179.8	7.0	9955	1	AE011700	AE011700 Xanthomon
33	172.8	6.7	30037	1	AE016927	AE016927 Bacteroid
34	170	6.6	303446	1	AE016941	AE016941 Bacteroid
35	155	6.0	299925	1	AP005045	AP005045 Streptomy
36	142.4	5.5	301761	1	AE016945	AE016945 Bacteroid
37	83.4	3.2	135552	2	AC019251	AC019251 Homo sapi
38	83	3.2	125020	2	AF429315	AF429315 Homo sapi
39	82.6	3.2	186752	9	AC111200	AC111200 Homo sapi
40	79.8	3.1	152686	2	AC143344	AC143344 Macaca mu
41	79.6	3.1	1094	11	PM7G11B	AL885196 Penicilliu
42	79.6	3.1	11305	8	OSAS35061	AJ535061 Oryza sat
43	79.6	3.1	159790	2	AP004786	AP004786 Oryza sat
44	79.4	3.1	65351	2	AC139773	AC139773 Homo sapi
45	79.2	3.1	69337	2	AC101261	AC101261 Mus muscu

## ALIGNMENTS

RESULT 1

LOCUS AR160672

DEFINITION Sequence 1 from patent US 6255096.

ACCESSION AR160672

VERSION AR160672.1 GI:16225095

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2575)

AUTHORS Hopwood, J., Joseph, Scott, H., Steele, Weber, B., Blanch, L. and Anson, D., Stewart.

TITLE Synthetic mammalian .alpha.-n-acetylglucosaminidase and genetic sequences encoding same

LOCUS AR160672 2575 bp DNA linear PAT 17-OCT-2001

JOURNAL Patent: US 6255096-A 1 03-JUL-2001;  
 FEATURES Location/Qualifiers  
 source 1..2575  
 /organism="unknown"  
 BASE COUNT 440 a 819 c 833 g 483 t  
 ORIGIN

Query Match 100.0%; Score 2575; DB 6; Length 2575;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CCCGGGCTTAAGCTTGGGGTCCAGTGGCCGGAGCCCGAGCTGATTGGACCCGGGCGG 60
Db 1 CCCGGGCTTAAGCTTGGGGTCCAGTGGCCGGAGCCCGAGCTGATTGGACCCGGGCGG 60
QY 61 CCCGACCCCTTGCGCGCTGCGGGAGCCCGACAGACTGAGACCATGAGAGCGGTGG 120
Db 61 CCCGACCCCTTGCGCGCTGCGGGAGCCCGACAGACTGAGACCATGAGAGCGGTGG 120
QY 121 CCGCGGCGGTGGGGGTCTTCTCTTGAGCCGGGGCCGGGAGCGGAGAGAGAGCC 180
Db 121 CCGCGGCGGTGGGGGTCTTCTCTTGAGCCGGGGCCGGGAGCGGAGAGAGAGCC 180
QY 181 GGGAGGCGGCGCGCTGCGGGCGCTGCGGCGCGGCTGCTGAGGCGAGGCCCGGCGG 240
Db 181 GGGAGGCGGCGCGCTGCGGGCGCTGCGGCGCGGCTGCTGAGGCGAGGCCCGGCGG 240
QY 241 ACTTCTCGGTGCTGGTGGAGCGGCTTGTGCTGCCAGACCGGAGCTTGGACACTTAC 300
Db 241 ACTTCTCGGTGCTGGTGGAGCGGCTTGTGCTGCCAGACCGGAGCTTGGACACTTAC 300
QY 301 TGGGCGGCGGCGGCGCGCGCGCTGCGGGTGGCGGGCTCAAGGGCGTGGCGGCGGCGG 360
Db 301 TGGGCGGCGGCGGCGCGCGCGCTGCGGGTGGCGGGCTCAAGGGCGTGGCGGCGGCGG 360
QY 361 CCGGGCTGCAACCGCTACCTGCGGACTTGTGGCTGCCACGTGGCGCTGGTCCGCTCTC 420
Db 361 CCGGGCTGCAACCGCTACCTGCGGACTTGTGGCTGCCACGTGGCGCTGGTCCGCTCTC 420
QY 421 AGCGGCGCGCTGCGGCGCGCGCTGCGGCGCGCTGCGGCGGAGCTGACCGAGCGCA 480
Db 421 AGCGGCGCGCTGCGGCGCGCGCTGCGGCGCGCTGCGGCGGAGCTGACCGAGCGCA 480
QY 481 ACAAGTACCGCTATTACAGAAATGTGTGCAAGCAAACTACTCTTCTGCTGGTGGAGT 540
Db 481 ACAAGTACCGCTATTACAGAAATGTGTGCAAGCAAACTACTCTTCTGCTGGTGGAGT 540
QY 541 GGGCGCGCTGGAGAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db 541 GGGCGCGCTGGAGAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 601 CTTGAGAGCGGCGAGAGAGCAATGAGAGCGGCTGATACCTGAGGCTTGGAGCCAGG 660
Db 601 CTTGAGAGCGGCGAGAGAGCAATGAGAGCGGCTGATACCTGAGGCTTGGAGCCAGG 660
QY 661 CAGAGATCAATGAGATTCTTACTGTGCTGCTCTTCTGCTGCGCTGGGGCGAATGGGCA 720
Db 661 CAGAGATCAATGAGATTCTTACTGTGCTGCTCTTCTGCTGCGCTGGGGCGAATGGGCA 720
QY 721 TGCACACCTGGAGATGGCCCCCTGCCCCCTCTGGAGCATCAAGAGAGCTTAACTGGAG 780
Db 721 TGCACACCTGGAGATGGCCCCCTGCCCCCTCTGGAGCATCAAGAGAGCTTAACTGGAG 780
QY 781 ACCGGGCTCTGACACAGATGCGCTCTTGGAGTGAACCCAGTGTCTGCTGATTTGGCG 840
Db 781 ACCGGGCTCTGACACAGATGCGCTCTTGGAGTGAACCCAGTGTCTGCTGATTTGGCG 840
QY 841 GGGATGTTCCCGAGGCTGTGACACAGAGGTGTTCCCTAGAGTCAATGTCAGAGAGTGGGA 900
Db 841 GGGATGTTCCCGAGGCTGTGACACAGAGGTGTTCCCTAGAGTCAATGTCAGAGAGTGGGA 900
QY 901 GTTGGGCGCACTTAACTGTCTACTCTGCTCTTCTCTTCTGCTGGAGAGAGCCCA 960
  
```

```

Db 901 GTTGGGCGCACTTAACTGTCTACTCTGCTCTTCTCTTCTGCTGGAGAGAGCCCA 960
QY 961 TATTCCTCATATGAGAGAGCTCTTCCGAGAGAGCTGATCAAGAGATTGGACAGAGC 1020
Db 961 TATTCCTCATATGAGAGAGCTCTTCCGAGAGAGCTGATCAAGAGATTGGACAGAGC 1020
QY 1021 ACATCTATGAGGCGGACACTTTCATATGATGAGAGCCACTTCTCTAGAGGCTCTTACC 1080
Db 1021 ACATCTATGAGGCGGACACTTTCATATGATGAGAGCCACTTCTCTAGAGGCTCTTACC 1080
QY 1081 TTGCGGAGAGCAACAGCTGCGCTATGAGAGGAGCATGAGAGTGAATATAGAGCTGAT 1140
Db 1081 TTGCGGAGAGCAACAGCTGCGCTATGAGAGGAGCATGAGAGTGAATATAGAGCTGAT 1140
QY 1141 GAGTCTTCAAGAGCTGCTCTTCCAGACACAGAGCGAGTCTGGGGGCGGCGCCAGATCA 1200
Db 1141 GAGTCTTCAAGAGCTGCTCTTCCAGACACAGAGCGAGTCTGGGGGCGGCGCCAGATCA 1200
QY 1201 GGGCTGTGCTGGAGAGCTGTGCTGCGGCGCTCTGCTGCTTCTGAGACTGTTTGTAGA 1260
Db 1201 GGGCTGTGCTGGAGAGCTGTGCTGCGGCGCTCTGCTGCTTCTGAGACTGTTTGTAGA 1260
QY 1261 GCCAGCTGTGTATACCGGCACTGCTCTTCCAGAGCGCAGCCCTTATCTGTGATGTC 1320
Db 1261 GCCAGCTGTGTATACCGGCACTGCTCTTCCAGAGCGCAGCCCTTATCTGTGATGTC 1320
QY 1321 TGCACAACTTTGGGGGAAACCATGCTCTTTTGGAGCCCTTAGAGGCTGTGAAACGAGGCC 1380
Db 1321 TGCACAACTTTGGGGGAAACCATGCTCTTTTGGAGCCCTTAGAGGCTGTGAAACGAGGCC 1380
QY 1381 CAGAAGCTGCGGCTCTTCCCAACTCAACATGATGAGGACGGGATGGAGCCCGGAGG 1440
Db 1381 CAGAAGCTGCGGCTCTTCCCAACTCAACATGATGAGGACGGGATGGAGCCCGGAGG 1440
QY 1441 GCATCAGCAGAAAGAAAGAGTGTATTCCTCATGAGCTGGGAGAAAGAGC 1500
Db 1441 GCATCAGCAGAAAGAAAGAGTGTATTCCTCATGAGCTGGGAGAAAGAGC 1500
QY 1501 CAGTGCAGATTGTGGAGCGCTGGGTGACAGCTTGTGCGCGCGGCTATGGGGTCTCC 1560
Db 1501 CAGTGCAGATTGTGGAGCGCTGGGTGACAGCTTGTGCGCGCGGCTATGGGGTCTCC 1560
QY 1561 ACCCGAGCGAGGGGAGCGTGAAGCTACTGCTCCGAGTGTGTACACTGCTCCGAGG 1620
Db 1561 ACCCGAGCGAGGGGAGCGTGAAGCTACTGCTCCGAGTGTGTACACTGCTCCGAGG 1620
QY 1621 AGGCTGTGAGGGGCGCAATGTGAGCCGCTGTGACAGCGCGCTCCCTACAGATGATA 1680
Db 1621 AGGCTGTGAGGGGCGCAATGTGAGCCGCTGTGACAGCGCGCTCCCTACAGATGATA 1680
QY 1681 CCAGCATCTGATCAACCGATCTGATGTGTTGAGGCTGTGGCGCTGCTCATCTG 1740
Db 1681 CCAGCATCTGATCAACCGATCTGATGTGTTGAGGCTGTGGCGCTGCTCATCTG 1740
QY 1741 CTCCCTCCCTGGGCGACAGCGCCGCTTCCGCTACAGCTGTGAGACTTCACTGAGAG 1800
Db 1741 CTCCCTCCCTGGGCGACAGCGCCGCTTCCGCTACAGCTGTGAGACTTCACTGAGAG 1800
QY 1801 CAGTGCAGAGAGCTGATGAGCTTGTACTATGAGAGGAGCAAGAGGCTTACTGAGCAAG 1860
Db 1801 CAGTGCAGAGAGCTGATGAGCTTGTACTATGAGAGGAGCAAGAGGCTTACTGAGCAAG 1860
QY 1861 AGTGGGCTCTCTGTTGAGAGGCTGAGAGGCTCTGAGCTTATGAGCTGCTGCGGACTGG 1920
Db 1861 AGTGGGCTCTCTGTTGAGAGGCTGAGAGGCTCTGAGCTTATGAGCTGCTGCGGACTGG 1920
QY 1921 ACGAGGTGTGGCTAGTGAAGAGCGGCTTGTGCTGGGCACTGAGCTGAGAGCGGCGAG 1980
Db 1921 ACGAGGTGTGGCTAGTGAAGAGCGGCTTGTGCTGGGCACTGAGCTGAGAGCGGCGAG 1980
QY 1981 CAGGCGAGTCAAGTGAAGCGGAGCGGCTTCTACAGAGCAAGACCGGCTACCACTGA 2040
Db 1981 CAGGCGAGTCAAGTGAAGCGGAGCGGCTTCTACAGAGCAAGACCGGCTACCACTGA 2040
  
```

QY 2041 CCTTGTGGGGCCAGAGGCAACATCTGGACTATGCAACAAGCAGCTGCGGGGTTGG 2100  
 Db 2041 CCTTGTGGGGGGCCAGAGGCAACATCTGGACTATGCAACAAGCAGCTGCGGGGTTGG 2100  
 QY 2101 TGGCCAACTACTACACCCCTGGCTGGGGGCTTTTCTGGAGGGCTGGTGAAGTGTGG 2160  
 Db 2101 TGGCCAACTACTACACCCCTGGCTGGGGGCTTTTCTGGAGGGCTGGTGAAGTGTGG 2160  
 QY 2161 CCCAGGCGCATCCCTTCCCAACAGCAGCAGTTGACAAAATGTCTTCAACTGAGCAGG 2220  
 Db 2161 CCCAGGCGCATCCCTTCCCAACAGCAGCAGTTGACAAAATGTCTTCAACTGAGCAGG 2220  
 QY 2221 CCTTGTCTTCAAGCAGAGGATACCCAGCCAGCCGCGAGAGACACTGTGAGCTGG 2280  
 Db 2221 CCTTGTCTTCAAGCAGAGGATACCCAGCCAGCCGCGAGAGACACTGTGAGCTGG 2280  
 QY 2281 CCAAGAGATCTTCCCAATATTTACCCGGGCTGGGTGGCCGGCTGTGTGATGATTC 2340  
 Db 2281 CCAAGAGATCTTCCCAATATTTACCCGGGCTGGGTGGCCGGCTGTGTGATGATTC 2340  
 QY 2341 GCCACCACTGGGCTTGTCTTCCGCTAATTCAGGGCAGATTCCAGGGCCAGAGCTGGA 2400  
 Db 2341 GCCACCACTGGGCTTGTCTTCCGCTAATTCAGGGCAGATTCCAGGGCCAGAGCTGGA 2400  
 QY 2401 CAGACATCAAGAGATTAACCCAGGCTGGAGAGAGGCCCCAGCGGCTGTGTGGGTCTG 2460  
 Db 2401 CAGACATCAAGAGATTAACCCAGGCTGGAGAGAGGCCCCAGCGGCTGTGTGGGTCTG 2460  
 QY 2461 ACCTGGGGGAGTTGAGAGGAAATGACCTGCTCCACACACCCAAAGTGTGGATTGA 2520  
 Db 2461 ACCTGGGGGAGTTGAGAGGAAATGACCTGCTCCACACACCCAAAGTGTGGATTGA 2520  
 QY 2521 AGTACTGTTTCTTCTTCACTTAATAAAAAAGTGAAGGCGCGGAATTC 2575  
 Db 2521 AGTACTGTTTCTTCTTCACTTAATAAAAAAGTGAAGGCGCGGAATTC 2575

RESULT 2  
 HSU40846 2575 bp mRNA linear pri 15-AUG-1996  
 LOCUS HSU40846 2575 bp mRNA linear pri 15-AUG-1996  
 DEFINITION Human alpha-N-acetylglucosaminidase (NAG) mRNA, complete cds.  
 VERSION U40846  
 ACCESSION U40846.1 GI:1197839  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 2575)  
 AUTHORS Weber,B., Blanch,L., Clements,P.R., Scott,H.S. and Hopwood,J.J.  
 TITLE Cloning and expression of the gene involved in Sanfilippo B syndrome (mucopolysaccharidosis III B)  
 JOURNAL Hum. Mol. Genet. 5 (6), 771-777 (1996)  
 MEDLINE 96372812  
 PUBMED 8776591  
 REFERENCE 2 (bases 1 to 2575)  
 AUTHORS Weber,B., Scott,H. and Hopwood,J.J.  
 TITLE Direct Submision  
 JOURNAL Submitted (17-NOV-1995) Birgit Weber, Chemical Pathology, Women's & Children's Hospital, King William Street 72, North Adelaide, S.A. 5006, Australia

FEATURES  
 source  
 1. .2575  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="117q12-21"  
 1. .2575  
 /gene="NAG"  
 102. .2333  
 /gene="NAG"

BASE COUNT 440 a 819 c 833 g 483 t  
 ORIGIN  
 Query Match 100.0%; Score 2575; DB 9; Length 2575;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CCCGGCTTACCTTGGGTCACGTGCGCGGAGCCGCGAGCTGATTGACGCGGCGG 60  
 1 CCCGGCTTACCTTGGGTCACGTGCGCGGAGCCGCGAGCTGATTGACGCGGCGG 60  
 61 CCCACCCCTGCGCGCTGCGGAGCCCGCAGAGTGAACCATGAGCGGTGGCGGTGG 120  
 61 CCCACCCCTGCGCGCTGCGGAGCCCGCAGAGTGAACCATGAGCGGTGGCGGTGG 120  
 61 CCCACCCCTGCGCGCTGCGGAGCCCGCAGAGTGAACCATGAGCGGTGGCGGTGG 120  
 121 CCGCGCGGTGGGGGTCTTCTCTGCGCGGCGCGGCGCGGCGCGGCGGCGGCGG 180  
 121 CCGCGCGGTGGGGGTCTTCTCTGCGCGGCGCGGCGCGGCGCGGCGGCGGCGG 180  
 121 CCGCGCGGTGGGGGTCTTCTCTGCGCGGCGCGGCGCGGCGCGGCGGCGGCGG 180  
 181 CCGAGCGCGCGCGCGCTGCGGCGCGCTGCGGCGCGGCGCGGCGCGGCGGCGG 240  
 181 CCGAGCGCGCGCGCGCTGCGGCGCGCTGCGGCGCGGCGCGGCGCGGCGGCGG 240  
 241 ACTTCTCGGTGCGGTGAGCGCGCTGCGTGCAGACCGCGGCTTGGACACCTACAGCC 300  
 241 ACTTCTCGGTGCGGTGAGCGCGCTGCGTGCAGACCGCGGCTTGGACACCTACAGCC 300  
 301 TGGCGCGCGCGCGCGCGCGCGCTGCGGCGCGGCGCTGCGGCGCGCGCGCGCG 360  
 301 TGGCGCGCGCGCGCGCGCGCGCTGCGGCGCGGCGCTGCGGCGCGCGCGCGCG 360  
 361 CCGGAGCTGCAACGCTTACCTGCGGCACTTCTGTGCTGCCAGTGGCCTGTGCTG 420  
 361 CCGGAGCTGCAACGCTTACCTGCGGCACTTCTGTGCTGCCAGTGGCCTGTGCTG 420  
 421 AGTGGCGCTGCGCGCGCGCGCTGCGGCGCGGCGGCGGCGGCGGCGGCGGCGG 480  
 421 AGTGGCGCTGCGCGCGCGCGCGCTGCGGCGCGGCGGCGGCGGCGGCGGCGG 480  
 481 ACAGGTACCGGTATTTACAGATGTGTCACGCGGAGGCTGCTTCTGTGTGTGACT 540  
 481 ACAGGTACCGGTATTTACAGATGTGTCACGCGGAGGCTGCTTCTGTGTGTGACT 540  
 541 GGGCGCGCTGCGGCGGAGATGATGATGAGTGGGCTGGAATGCGATCAACTGCACTG 600  
 541 GGGCGCGCTGCGGCGGAGATGATGATGAGTGGGCTGGAATGCGATCAACTGCACTG 600  
 601 CCGGAGCGCGCGGAGGCGCATGCGAGCGGCGGCTGCGGCGGCGGCGGCGGCGG 660  
 601 CCGGAGCGCGCGGAGGCGCATGCGAGCGGCGGCTGCGGCGGCGGCGGCGGCGG 660  
 661 CAGAGATCAATGATTTACTGTGCTGCTGCTGCGGCGGCGGCGGAGTGGGCAACC 720

Db	661	CAGAGATCAATGAGTCTTTACTGTGTCGTGCTCTCTGAGCCTGAGGAGCGAATGAGCAAC	722
QY	721	TGCACACTGGGATGAGCCCCCTGCCCCCTCTGTGCATATCAAGTACTTTACTGTCAAC	780
Db	721	TGCACACTGGGATGAGCCCCCTGCCCCCTCTGTGCATATCAAGTACTTTACTGTCAAC	780
QY	781	ACCAGGTCCTGAGACAGATGAGGCTCTTCGTGGCATACCCAGTGTGCTCATTCGCG	840
Db	781	ACCAGGTCCTGAGACAGATGAGGCTCTTCGTGGCATACCCAGTGTGCTCATTCGCG	840
QY	841	GGCATGTTCCGAGGCTGTCAACAGAGGTTCCTCAAGTCAATGTCAAGAAATGAGCA	900
Db	841	GGCATGTTCCGAGGCTGTCAACAGAGGTTCCTCAAGTCAATGTCAAGAAATGAGCA	900
QY	901	GTTGGAGGCACATTTACTGTCTTACTCTGTCTGTCTCTTCCTTCCTTCGTGCTCGGAAAGCCCCA	960
Db	901	GTTGGAGGCACATTTACTGTCTTACTCTGTCTGTCTCTTCCTTCCTTCGTGCTCGGAAAGCCCCA	960
QY	961	TATTCGCCCATCATCGGAGACCTCTTCGCGAGAGCTGATCAAAAGTTTGACACAGAC	1020
Db	961	TATTCGCCCATCATCGGAGACCTCTTCGCGAGAGCTGATCAAAAGTTTGACACAGAC	1020
QY	1021	ACATCTAATGGGCGCAGCTTTGCATGAGATGACAGCCACTTCTCCACAGCCCTCTAC	1080
Db	1021	ACATCTAATGGGCGCAGCTTTGCATGAGATGACAGCCACTTCTCCACAGCCCTCTAC	1080
QY	1081	TTGCCGAGCAGCACCACTGCGGCTATGAGGCGCATATGCAAGTGAATACTAGAGCTGAT	1140
Db	1081	TTGCCGAGCAGCACCACTGCGGCTATGAGGCGCATATGCAAGTGAATACTAGAGCTGAT	1140
QY	1141	GCGTGTCTCMAAGGCTGTCTTTCCAGCACACAGCCGAGTTCTGGGAGCCCGCCAGATCA	1200
Db	1141	GCGTGTCTCMAAGGCTGTCTTTCCAGCACACAGCCGAGTTCTGGGAGCCCGCCAGATCA	1200
QY	1201	GAGGCTGTCTGAGAGCTGTGACCCGCGGTGAGCGCGCTGTGATTCGACCTGTGTTGTGAGA	1260
Db	1201	GAGGCTGTCTGAGAGCTGTGACCCGCGGTGAGCGCGCTGTGATTCGACCTGTGTTGTGAGA	1260
QY	1261	GCCAGCCTGTGTATACCCGCACTGTCTCTTCAGAGGCAAGCCCTTCACTTGTGTGATGC	1320
Db	1261	GCCAGCCTGTGTATACCCGCACTGTCTCTTCAGAGGCAAGCCCTTCACTTGTGTGATGC	1320
QY	1321	TGCACAACTTTGGGGGAAACCATGTCTTTTGGAGCCCTGAGGCTGTGAACGAGAGCC	1380
Db	1321	TGCACAACTTTGGGGGAAACCATGTCTTTTGGAGCCCTGAGGCTGTGAACGAGAGCC	1380
QY	1381	CAGAACTGTCCGCTCTTCCCACTCCACCATGTATGAGGACAGGCGCATGAGCCGCCAGG	1440
Db	1381	CAGAACTGTCCGCTCTTCCCACTCCACCATGTATGAGGACAGGCGCATGAGCCGCCAGG	1440
QY	1441	GCATATGCGCAAGACGAATGGTCTATTTCCCTCATATGCTGAGCTGGGCTGGCGAAAGAAC	1500
Db	1441	GCATATGCGCAAGACGAATGGTCTATTTCCCTCATATGCTGAGCTGGGCTGGCGAAAGAAC	1500
QY	1501	CAGTCCCAAGTTTGGCAGGCTGTGGAGGACAGGCTTTGACCGCCGAGCGATGAGGCTCC	1560
Db	1501	CAGTCCCAAGTTTGGCAGGCTGTGGAGGACAGGCTTTGACCGCCGAGCGATGAGGCTCC	1560
QY	1561	ACCAGGACGACAGGAGCAGCTGTGAGAGCTATCTGCTCCGAGATGTGTACAACTGTCCGAGG	1620
Db	1561	ACCAGGACGACAGGAGCAGCTGTGAGAGCTATCTGCTCCGAGATGTGTACAACTGTCCGAGG	1620
QY	1621	AGGCTGTGAGGGCCACATGCTTACCCGCTGTGACAGCGCCGTCCTTACAGATGATA	1680
Db	1621	AGGCTGTGAGGGCCACATGCTTACCCGCTGTGACAGCGCCGTCCTTACAGATGATA	1680
QY	1681	CCAGCATCTGTGACAAACGATCTGATGTGTGAGGCTGTGAGGCTGTGCTCATCATCTG	1740
Db	1681	CCAGCATCTGTGACAAACGATCTGATGTGTGAGGCTGTGAGGCTGTGCTCATCATCTG	1740
QY	1741	CTCCCTCCCTGTGACACAGCCCGGCTTCCGCTATGACCTGTGTGACTCACTGGGACAG	1800
Db	1741	CTCCCTCCCTGTGACACAGCCCGGCTTCCGCTATGACCTGTGTGACTCACTGGGACAG	1800

[illegible]







QY	396	TGCACAGTGGCTGGATCCGGCTCTGAGCTGGCCCTGGCCGACATCCAGCCCTGGC	455
Db	301	TGCCAGTGGCTGGATCCGGCTCTGAGCTGGCCCTGGCCGACATCCAGCCCTGGC	360
QY	456	GCGGAGCTGACCGGAGCCACCGCCCAACAGTACCGCTATTACAGAAATGTGTGACGCGAA	515
Db	351	GCGGAGCTGACCGGAGCCACCGCCCAACAGTACCGCTATTACAGAAATGTGTGACGCGAA	420
QY	516	AGCTACTCTCTTGTGTGTGGGACCTGGGCGCGCTGGGAGCGAGATAGATCTGATGGACG	575
Db	421	AGCTACTCTCTTGTGTGTGGGACCTGGGCGCGCTGGGAGCGAGATAGATCTGATGGACG	480
QY	576	CTGAATGCGATCAACCTTGCGCATGGCCCTGGAGCGGACCGAGAGCCATCTGGACGGGTG	635
Db	481	CTGAATGCGATCAACCTTGCGCATGGCCCTGGAGCGGACCGAGAGCCATCTGGAGGGGTG	540
QY	636	TACCTAGGCTTTGGGCTGTGACCGAGGACAGATCATAGATCTTTACTGTGTCTGGCTTC	695
Db	541	TACCTAGGCTTTGGGCTGTGACCGAGGACAGATCATAGATCTTTACTGTGTCTGGCTTC	600
QY	656	CTGGCCTTGGGGGCGAATGGGCAACCTTGCAACCTTGGAATGGCCCTGCCCCCTCTCGG	755
Db	601	CTGGCCTTGGGGGCGAATGGGCAACCTTGCAACCTTGGAATGGCCCTGCCCCCTCTCGG	660
QY	756	CACATCAAGCAGCTTTACTCTTGAGACCGGGGCTCGGACAGATGGGCTCTTGGCATG	815
Db	661	CACATCAAGCAGCTTTACTCTTGAGACCGGGGCTCGGACAGATGGGCTCTTGGCATG	720
QY	816	ACCCAGTGTCTGCTGCATTTGGCGGGCATGTTCCGAGGCTGTGACCAAGGATGTTCCCT	875
Db	721	ACCCAGTGTCTGCTGCATTTGGCGGGCATGTTCCGAGGCTGTGACCAAGGATGTTCCCT	780
QY	876	CAGGTCAATGTGACGAGATGGGGAGTGGGGCCACTTAATGTTCTTACTCTCCGCTCC	935
Db	781	CAGGTCAATGTGACGAGATGGGGAGTGGGGCCACTTAATGTTCTTACTCTCCGCTCC	840
QY	936	TTCTCTTGGCTCCGGAAGACCCCATATTCCCATCATCTGGGAGCCCTTCTCTGGAGAG	995
Db	841	TTCTCTTGGCTCCGGAAGACCCCATATTCCCATCATCTGGGAGCCCTTCTCTGGAGAG	900
QY	996	CTGATCAAGAGTTTGGGACAGACCAACATCTATGGGGCGGACACTTAAATGAGATGGAG	1055
Db	901	CTGATCAAGAGTTTGGGACAGACCAACATCTATGGGGCGGACACTTAAATGAGATGGAG	960
QY	1056	CCACCTTCTCAGAGCCCTCTACCTTCCGCGAGCCACACTGCCGTATGAGGCCATG	1115
Db	961	CCACCTTCTCAGAGCCCTCTACCTTCCGCGAGCCACACTGCCGTATGAGGCCATG	1020
QY	1116	ACTGAGTGGATCTGAGGCTGTGTGGCTGGCTCCAGAGCTGGCTTCCACACCAAGCG	1175
Db	1021	ACTGAGTGGATCTGAGGCTGTGTGGCTGGCTCCAGAGCTGGCTTCCACACCAAGCG	1080
QY	1176	CAGTTCTGGGGGCCCGCCAGATCAAGGCTGTGTGGAGCTGTGGTCCCGTGGCCGCGCTC	1235
Db	1081	CAGTTCTGGGGGCCCGCCAGATCAAGGCTGTGTGGAGCTGTGGTCCCGTGGCCGCGCTC	1140
QY	1236	CTGGTCTTGAACCTGTTGTGTGAGAGGACAGCTGTGTATACCGGACCTGCTCCCTCCAG	1295
Db	1141	CTGGTCTTGAACCTGTTGTGTGAGAGGACAGCTGTGTATACCGGACCTGCTCCCTCCAG	1200
QY	1296	GCGCAGCCCTTCATCTGGTGAATGCTGCAAACTTTGGGGGAAACATGATGTTTGTGGA	1355
Db	1201	GCGCAGCCCTTCATCTGGTGAATGCTGCAAACTTTGGGGGAAACATGATGTTTGTGGA	1260
QY	1356	GCCCTAAGGCTGTGAAACGAGAGCCCAAGAGTGCCCGCTCTTCCCAATCCACACATG	1415
Db	1261	GCCCTAAGGCTGTGAAACGAGAGCCCAAGAGTGCCCGCTCTTCCCAATCCACACATG	1320
QY	1416	GTAAGGACGGGCATGGCCCCCGAGGGCATCAGCCAGAAAGATGTCTATTCCTCATG	1475
Db	1321	GTAAGGACGGGCATGGCCCCCGAGGGCATCAGCCAGAAAGATGTCTATTCCTCATG	1380

QY	1476	GCTAGCTGGGCTGGCCAAAGAACCCAGCTGCACATTTGGAGCGCTGGTGACAGACTT	1535
Db	1381	GCTAGCTGGGCTGGCCAAAGAACCCAGCTGCACATTTGGAGCGCTGGTGACAGACTT	1440
QY	1536	GCCGCCCGCGGTATGGGGTCTCCACCCGAGCCAGGGCAGCGTGGAGGCTACTGTC	1595
Db	1441	GCCGCCCGCGGTATGGGGTCTCCACCCGAGCCAGGGCAGCGTGGAGGCTACTGTC	1500
QY	1596	CGAGTGTGTACAATGCTCCGGGGAGGCGCTGCAGAGGGGCCAATTCGTAGCCGCTGGTC	1655
Db	1501	CGAGTGTGTACAATGCTCCGGGGAGGCGCTGCAGAGGGGCCAATTCGTAGCCGCTGGTC	1560
QY	1656	AGCGGGCCGTCCTCTACAGATGAATACGACATCTGTACAACCGATCTGATGTGTTGAG	1715
Db	1561	AGCGGGCCGTCCTCTACAGATGAATACGACATCTGTACAACCGATCTGATGTGTTGAG	1620
QY	1716	GCTGGCGGCGCTGCTGCATCTGTGCTCCCTCCCTGGGACACAGCCCCGCTTCGCGTAC	1775
Db	1621	GCTGGCGGCGCTGCTGCATCTGTGCTCCCTCCCTGGGACACAGCCCCGCTTCGCGTAC	1680
QY	1776	GACCTGTGACCTCACTCGGCAAGTGCAGAGAGCTGTCAGCTTGATCTATAGAGAG	1835
Db	1681	GACCTGTGACCTCACTCGGCAAGTGCAGAGAGCTGTCAGCTTGATCTATAGAGAG	1740
QY	1836	GCAAGAAAGCGCTTACTGTAGCAAGAGCTGGCCCTCCGTGTAAGGGCTGAGGCGTCTG	1895
Db	1741	GCAAGAAAGCGCTTACTGTAGCAAGAGTGGCCCTCCGTGTAAGGGCTGAGGCGTCTG	1800
QY	1896	GCTATAGCTGCTGTCGCGCACTGGAAGAGTGTGCTAGTGAACAGCCGCTTCTTGCTG	1955
Db	1801	GCTATAGCTGCTGTCGCGCACTGGAAGAGTGTGCTAGTGAACAGCCGCTTCTTGCTG	1860
QY	1956	GGCAGCTGGCTAGAGCAGAGCCCGAGCAGCCGCACTAGTGAAGCCGAGGCCGATTCTAC	2015
Db	1861	GGCAGCTGGCTAGAGCAGAGCCCGAGCAGCCGCACTAGTGAAGCCGAGGCCGATTCTAC	1920
QY	2016	GAGCAGAAACAGCCGCTACCAAGCTGACTTGTGGGGGCCAAGAGGCAATCTCGGACTAT	2075
Db	1921	GAGCAGAAACAGCCGCTACCAAGCTGACTTGTGGGGGCCAAGAGGCAATCTCGGACTAT	1980
QY	2076	GCCAAACAAGCAGCTGGGGGGTGGTGGCCAACTACTACACCCCTGCTGGCGCTTTC	2135
Db	1981	GCCAAACAAGCAGCTGGGGGGTGGTGGCCAACTACTACACCCCTGCTGGCGCTTTC	2040
QY	2136	CTGAGAGCGCTGTGTGACAGTGTGGCCCAAGGGCATCCCTTTCACACACACAGATTGAC	2195
Db	2041	CTGAGAGCGCTGTGTGACAGTGTGGCCCAAGGGCATCCCTTTCACACACACAGATTGAC	2100
QY	2196	AAAAATGCTTCCAACTGAGCAGAGGCTTGGTTCAGCAAGCAGAGTACCCAGCCAG	2255
Db	2101	AAAAATGCTTCCAACTGAGCAGAGGCTTGGTTCAGCAAGCAGAGTACCCAGCCAG	2160
QY	2256	CCGCGAGAGACACTGTGGAACCTGGCCAAAGAGTCTTCTCAATATTAACCCCGCTGG	2315
Db	2161	CCGCGAGAGAGCACTGTGGAACCTGGCCAAAGAGTCTTCTCAATATTAACCCCGCTGG	2220
QY	2316	GTGGCCGGGCTTGGTATAGATTGCGCAACACTGGGCTTGTTTCCGCTAATTCAG	2375
Db	2221	GTGGCCGGGCTTGGTATAGATTGCGCAACACTGGGCTTGTTTCCGCTAATTCAG	2280
QY	2376	GCAGATTCAGGGCCCAAGCTGGAACAATCAAGATTAACCCAGGCTGGAGGAG	2435
Db	2281	GCAGATTCAGGGCCCAAGCTGGAACAATCAAGATTAACCCAGGCTGGAGGAG	2340
QY	2436	CCCCACGGCTGCTGTGGGGGTCTGACCTGGGGGATTTGAGGGAAATGACTGCTCC	2495
Db	2341	CCCCACGGCTGCTGTGGGGGTCTGACCTGGGGGATTTGAGGGAAATGACTGCTCC	2400
QY	2496	ACCAACCAACCAAGTGGGATTAAGTACTGTTTCTTTCACATT	2551
Db	2401	ACCAACCAACCAAGTGGGATTAAGTACTGTTTCTTTCACATT	2446

[illegible]

Db	996	GGAGCGGCACGAGGCGCATCTGGAGCGGGGTGTAACTGGCGCTTGAGGCGCTCAACCCAGGCGAG	1055
QY	664	AGATCAATGAGTTCTTTACTGAGTCTCCGCTTCCCTGGGCTGAGGGGGCGAAATGGGACAACTGC	723
Db	1056	AGATCAATGAGTTCTTTACTGAGTCTCCGCTTCCCTGGGCTGAGGGGGCGAAATGGGACAACTGC	1115
QY	724	ACACTGGGATGGCGCCCTGCGCCCTCTCGGACATCAAGCAAGCTTTTACTCTGACGACC	783
Db	1116	ACACTGGGATGGCGCCCTGCGCCCTCTCGGACATCAAGCAAGCTTTTACTCTGACGACC	1175
QY	784	GGGTCTTGAGACAGATGCGCTCTTGGGCAATGACCCCAATGCTGCTCTGCAATTGGCGGGGCG	843
Db	1176	GGGTCTTGAGACAGATGCGCTCTTGGGCAATGACCCCAATGCTGCTCTGCAATTGGCGGGGCG	1235
QY	844	ATGTTCGCGAGGCTGTACCAAGGCTGTTCCTCAAGTCAATGTGACAGAAATGGGACGTT	903
Db	1236	ATGTTCGCGAGGCTGTGTACCAAGGATGTTCCTCAAGTCAATGTGACAGAAATGGGACGTT	1295
QY	904	GGGCGCACTTTAACTGTCTCTACTCTGCTCTCTCTTCTTGCTCTGGAGAGACCCCAAT	963
Db	1296	GGGCGCACTTTAACTGTCTCTACTCTGCTCTCTCTTCTTGCTCTGGAGAGACCCCAAT	1355
QY	964	TCCCAATCATGGGAGGCGCTCTCTCGGAGAGAGTGAATCAAGAATTTGGACAGACACA	1023
Db	1356	TCCCAATCATGGGAGGCGCTCTCTCGGAGAGCTGAATCAAGAATTTGGACAGACACA	1415
QY	1024	TCTATGGGGCGGACACTTTCAATGAGATGAGCCACACTTCTCGAGCGCTCTTACTTG	1083
Db	1416	TCTATGGGGCGGACACTTTCAATGAGATGAGCCACACTTCTCTGAGCGCTCTTACTTG	1475
QY	1084	CCGAGACACACACTGCGCGCTCTATATAGAGCCATGATGAGATGAGATCAAGGCTGTGGC	1143
Db	1476	CCGAGACACACACTGCGCGCTCTATATAGAGCCATGATGAGATGAGAGCTGTGTGGC	1535
QY	1144	TGCTTCAAGGCTGAGCTTTCCAGACACAGCGCGAGTTCTGGGGCGCCGACCCAGATCAAGG	1203
Db	1536	TGCTTCAAGGCTGAGCTTTCCAGACACAGCGCGAGTTCTGGGGCGCCGACCCAGATCAAGG	1595
QY	1204	CTGTGCTGGAGCTGTGTGCCCCGTGGCGGCTCTCTGTGTTCTGGAACCTGTTTGTGAGAGCC	1253
Db	1596	CTGTGCTGGAGCTGTGTGCCCCGTGGCGGCTCTCTGTGTTCTGGAACCTGTTTGTGAGAGCC	1655
QY	1264	AGCGTGTATTAACCGGACACTGCTCTCTTCCAGAGGACAGACCCTTCAATGTGATGATGTC	1323
Db	1656	AGCGTGTATTAACCGGACACTGCTCTCTTCCAGAGGACAGACCCTTCAATGTGATGATGTC	1715
QY	1324	ACAACTTTGGGGGAAACCAATGATCTTTTGGAGCTTAGAGGCTGTGATCGGAGCCGAG	1383
Db	1716	ACAACTTTGGGGGAAACCAATGATCTTTTGGAGCTTAGAGGCTGTGATCGGAGCCGAG	1775
QY	1384	AAGCTGCCCGCTCTTCCCAACCTCCACCAATGTGATGAGGACAAGGAGATGGGCCCCGAGGCA	1443
Db	1776	AAGCTGCCCGCTCTTCCCAACCTCCACCAATGTGATGAGGACAAGGAGATGGGCCCCGAGGCA	1835
QY	1444	TCAGCCAGAAAGAAAGTGTCTATTCCCTCATGCTGAGCTGAGGCTGGCGGAAAGACCCAG	1503
Db	1836	TCAGCCAGAAAGAAAGTGTGTATTCCCTCATGCTGAGCTGAGGCTGGCGGAAAGACCCAG	1895
QY	1504	TGCCAGATTTGGGACGCTCGGGTGAACAAGCTTTGCCGCCCGCGGGTATGGGGTCTTCCACC	1563
Db	1896	TGCCAGATTTTGGGACGCTCGGGTGAACAAGCTTTGCCGCCCGCGGGTATGGGGTCTTCCACC	1955
QY	1564	CGGACGACAGGGGACGCTGAGAGCTTACTGCTCCGAGAGTGTATCAACTGCTCCGCGGAGG	1623
Db	1956	CGGACGACAGGGGACGCTGAGAGCTTACTGCTCCGAGAGTGTATCAACTGCTCCGCGGAGG	2015
QY	1624	CTTGAGGGGGCCAAATGTATGCCCCGTGATCAAGCGGCGCGTCTCTACATGATGATATCA	1683
Db	2016	CTTGAGGGGGCCAAATGTATGCCCCGTGATCAAGCGGCGCGTCTCTACATGATGATATCA	2075
QY	1684	GCATCTGGTATCAACCAATCTGATATGTTTGAAGCTGGCGGCTCTGCTCATCTGTCT	1743
Db	2076	GCATCTGGTATCAACCAATCTGATATGTTTGAAGCTGGCGGCTCTGCTCATCTGTCT	2135

QY 1744 CCTCCCTGGCCACAGCCGCCCTTCCGCTACGACCTGAGCTTCACTCGGAGGCGAG 1803  
 Db 2136 CCTCCCTGGCCACAGCCGCCCTTCCGCTACGACCTGAGCTTCACTCGGAGGCGAG 2195  
 QY 1804 TGCAGAGAGCTGTGAGCTTGTATATAGAGAGGCAAGAGCGCTACCTGAGCAAGAGC 1863  
 Db 2196 TGCAGAGAGCTGTGAGCTTGTATATAGAGAGGCAAGAGCGCTACCTGAGCAAGAGC 2255  
 QY 1864 TGGCTTCCCTGTTGAGGAGCTGAGAGCGCTCTGAGCTATGAGAGCTGAGCGGAGC 1923  
 Db 2256 TGGCTTCCCTGTTGAGGAGCTGAGAGCGCTCTGAGCTATGAGAGCTGAGCGGAGC 2315  
 QY 1924 AGGAGCTGAGTGAAGAGAGCGCTTCTGCTGAGAGCTGAGAGAGCGGAGCGAG 1983  
 Db 2316 AGGAGCTGAGTGAAGAGAGCGCTTCTGCTGAGAGCTGAGAGAGCGGAGCGAG 2375  
 QY 1984 CGGAGCTGAGTGAAGAGAGCGGATTTCTAGAGAGAGAGAGCGGAGCTGAGAGCT 2043  
 Db 2376 CGGAGCTGAGTGAAGAGAGCGGATTTCTAGAGAGAGAGAGCGGAGCTGAGAGCT 2435  
 QY 2044 TGTGGGGGGCCAGAGAGCAATCTGAGCTATGCTCAACAGAGCTGAGGAGGCTTGTG 2103  
 Db 2436 TGTGGGGGGCCAGAGAGCAATCTGAGCTATGCTCAACAGAGCTGAGGAGGCTTGTG 2495  
 QY 2104 CCAACTACTACACCCCTCGCTGAGCGCTTCTCTGAGAGCGCTGTTGAAGTGTGAGCC 2163  
 Db 2496 CCAACTACTACACCCCTCGCTGAGCGCTTCTCTGAGAGCGCTGTTGAAGTGTGAGCC 2555  
 QY 2164 AGGAGCTGCTTTTCCAGAGAGCAAGTGTGAGCAAAATGTCTTCAACTGAGAGAGGCT 2223  
 Db 2556 AGGAGCTGCTTTTCCAGAGAGCAAGTGTGAGCAAAATGTCTTCAACTGAGAGAGGCT 2615  
 QY 2224 TCGTCTCAGACAGAGAGGATCCGAGCGGAGCGGAGAGAGCACTGTGAGCTTGTGAG 2283  
 Db 2616 TCGTCTCAGACAGAGAGGATCCGAGCGGAGCGGAGAGAGCACTGTGAGCTTGTGAG 2675  
 QY 2284 AGAAGATCTTCTCAATATTTAACCCTGAGTGGCTGAGCTTGTGATGATTTGAGCC 2343  
 Db 2676 AGAAGATCTTCTCAATATTTAACCCTGAGTGGCTGAGCTTGTGATGATTTGAGCC 2735  
 QY 2344 ACCACTGAGGCTTGTGTTTCCGCTAATTTCCAGAGGAGAGATTCAGAGGCGGAGCTGAGCAG 2403  
 Db 2736 ACCACTGAGGCTTGTGTTTCCGCTAATTTCCAGAGGAGAGATTCAGAGGCGGAGCTGAGCAG 2795  
 QY 2404 ACATCAAGAGATTAACCAAGGCTGAGAGAGGCGGAGCGGCTGAGGAGGAGCTGAGCAG 2463  
 Db 2796 ACATCAAGAGATTAACCAAGGCTGAGAGAGGCGGAGCGGCTGAGGAGGAGCTGAGCAG 2855  
 QY 2464 TGGGGGAGATTGAGGAGAAATGACCTGCTCCACCAAGCTGAGGATTAAGT 2523  
 Db 2856 TGGGGGAGATTGAGGAGAAATGACCTGCTCCACCAAGCTGAGGATTAAGT 2915  
 QY 2524 ACTGTTTCTTTCACACT 2541  
 Db 2916 ACTGTTTCTTTCACACT 2933

RESULT 6  
 MMU85247 2452 bp mRNA linear ROD 03-DEC-1997  
 LOCUS complete cds.  
 DEFINITION Mus musculus lysosomal alpha-N-acetylglucosaminidase Naglu mRNA.  
 ACCESSION U85247  
 VERSION U85247.1 GI:2660687  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 2452)  
 AUTHORS Zhao,K.W., Li,H.H. and Neufeld,E.F.  
 TITLE Cloning and expression of Mouse gene encoding the lysosomal

JOURNAL  
 Unpublished  
 REFERENCE  
 2 (bases 1 to 2452)  
 AUTHORS  
 Zhao,K.W., Li,H.H. and Neufeld,E.F.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (13-JUN-1997) Biol. Chem., UCLA, 10833 Le Conte Ave, Los Angeles, CA 90095 USA  
 FEATURES  
 source  
 1..2452  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="B6/CDAP1J"  
 /db\_xref="taxon:10090"  
 /tissue\_type="Liver"  
 1..2220  
 /EC\_number="3.2.1.50"  
 /function="hydrolyse alpha-N-acetylglucosaminide from glycosaminoglycans"  
 /note="lysosomal alpha-N-acetylglucosaminidase"  
 /codon\_start=1  
 /product="Naglu"  
 /protein\_id="AAB88084.1"  
 /db\_xref="gi:2660688"  
 /translation="MEAGLAVALIGFLLAGSVDEAREKAVRELVRLGPAPAA  
 NFLVSVERALADESGDITVSLGGGQVPLVVRGSGTGVAAAAGLRVLRDPCGVAMS  
 SNOHLTPMLPAPVDPGLTETTPNRYRYVNVCHSYSPFWMDMARMEQIDMNLNGI  
 NLALANGQELINRWVTLALGDTSEIDITFTGAFPLAMGRMLHTMDGLPRLSHL  
 SQVYLQHRILDRKMSFGMITPLVPLAFACHVPAITRVFPVNVNLGSGMGHFNCSYCS  
 FLAPGDMPPLIGNFLRELTKFEGDHIYADTFENMOPFSPSPYLAATTAAYE  
 AMVAVDDPAVLLQGMLEFQHPQFQSPQIRAVLEAVPRRLVLVDLFLASHPYVMHT  
 ASFHGPFIMCMLEHFGNGLFGALDENVRGPQAAIPNSTVVGIGLPEIGONE  
 VVVALMELGMRDQPEVDLMAMWSSFAIRRYGSPDPAVAMKILLSVNSCEAS  
 GHNSPLVKRPSLSMSTAVYNSDVEARMLILTAAPNLTSPAFRYDLDITROY  
 OELVSLTSEARLTKQKEDLLIRAGGLVYKLPFLDELSSSFLGLGTWLDQAR  
 KAAVSEAAQFYEQNSRYQITLWGFENGLIDYANKQLAGVADYVQPRKCLFLGTDLH  
 SLANGVFPQDHEKKNVPLEQAFVNNKKEYPSPQDITVLSKRIPLKHPDPSL"  
 <1..377  
 exon  
 378..525  
 exon  
 526..672  
 exon  
 673..758  
 exon  
 759..1015  
 exon  
 1016..>2220  
 BASE COUNT  
 486 a 735 c 709 g 522 t  
 ORIGIN

Query Match 63.1%; Score 1623.6; DB 10; Length 2452;  
 Best Local Similarity 82.8%; Pred. No. 4; Le-207;  
 Matches 1864; Conservative 0; Mismatches 380; Indels 6; Gaps 1;

QY 110 GGTGGCGGTGGCCCGCGCGGTGGGGGTCTTCTCTCTGAGCGGAGCGGAGGCGGAGG 169  
 Db 3 GGAAGCGCGGGAGGTGGCCCGTGAATCTGGGGTTCCTACTCTGAGCTGGGGGCTGTGGG 62  
 QY 170 CGACGAGGCCCGGAGAGCGCGCGCTGCGGCGGCTGTGGCCCGGCTGCTGGGCGCAG 229  
 Db 63 TGATGAGGCTCGGAGGCAAGAGCGCGGAGAACTGTGTGCTGCTGAGTGGACCCGG 122  
 QY 230 CCCCAGCGCACTTCTCGGTGTGAGTGAAGCGGCTGTGAGCTTCTGAGGCTGAGGCGCTG 289  
 Db 123 GCCGCGCGCAATTTCTTGTGTCTGAGAGCGGCGCTGTGGAGAGAGTCTGGGCTGGA 182  
 QY 290 CACTTAAGCTTGGCGCGCGCGCGCGCGCGCGCTGCGGAGTGTGGCTTCCAGCGGCGT 349  
 Db 183 CACTTAAGCTTGGCGCGCGCGCGCGCGCGCGCGAGTGTGTGCTGCGCGGCTCCAGCGGCGT 242  
 QY 350 GGGCGCGCGCGGGGTGACCGCTACCTGCGGAGATTTCTGAGGCTGAGCGGCTGAGGCTG 409  
 Db 243 GCGGCGAGCGCGGGCTGACCGCTACCTGCTGAGCTTGTGTGCTGAGTGTGAGGCTG 302  
 QY 410 GTCCGAGCTTCACTGAGCTTGGCGCGCGCGCGCGCGCGCTGCGAGCGGAGGCTGACCGA 469  
 Db 303 GTCCAGGCTCACTGAGCTTGGCGCGCGCGCGCGCGCGCTGCTGTGCTGCGCGAGCGGCTGACCGA 362

470 GGCCAGCCCAAGATACCGCTATTAACAAGATGTGACAGCAAGTACTCTTCGT 529  
422 AACCAACCCCAAGATACCGCTATTAACAAGATGTGACAGCAAGTACTCTTCGT 422  
530 GTGGTGGGACTGGGCGGCTGGAGGAGAGATAGACTGATGGCGGTGAATGGATCAA 589  
423 GTGGTGGGACTGGGCGGCTGGAGGAGATAGACTGATGGCGGTGAATGGATCAA 482  
590 CCTGGCACTGGGCGGCTGGAGGAGGAGATAGACTGATGGCGGTGAATGGATCAA 649  
483 CTGGCTCTGGCTTGAATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 542  
650 CCTGACCCAGGAGAGATCAATGAGTCTTTAATGCTGCTGCTGCTGCTGCTGCTGCT 709  
543 CCTGACCCAGGAGAGATCAATGAGTCTTTAATGCTGCTGCTGCTGCTGCTGCTGCT 602  
710 AATGGGCACTGACACATGGATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 769  
603 CATGGTAACTTGAACACCTGGAGTGGCGGCTGGCGGCTGGCGGCTGGCGGCTGAGT 662  
770 TTACCTGACGACCGGGTCTGACACAGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 829  
663 CTACCTGACGACCGGGTCTGACACAGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 722  
830 TGCAATGCGGGGAGATGTTCCCGAGGCTGCAACAGGCTGTTCCCTGAGGCTCATGTGAC 889  
723 TGCCCTGCGAGGGAGATGCTCCCAAGGCGCATCACAGGCTGTTCCCAAGGCTCATGTGAC 782  
890 GAAGATGGGAGTGGGGGCACTTTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 949  
783 CAATGGGGAGCTGGGGACATTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 842  
950 GGAAGACCCCAATTTCCCATCATGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1009  
843 AGGAGACCCCAATTTCCCATCATGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 902  
1010 TGCGACAGACCATCTATGGGGGCGGACATTTCAATGAGATGAGGACATTTCTCTCAAG 1069  
903 TGCGACAGACCATCTATGGGGGCGGACATTTCAATGAGATGAGGACATTTCTCTCTGGA 962  
1070 GGCCTCTCACTTCCCGAGGACACACTGCGCTGCTATGAGGCGCATGCTGCTGCTGCTGCT 1129  
963 GGCCTCTCACTTCCCGAGGACACACTGCGCTGCTATGAGGCGCATGCTGCTGCTGCTGCT 1022  
1130 TGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1189  
1023 TGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1082  
1190 CGCCCAAGATCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1249  
1083 CTCTCAATCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1142  
1250 GTTTCCTGAGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1309  
1143 GTTTCCTGAGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1202  
1310 CTGGTGCATGCTGCAACTTTGGGGGAAACATGCTCTTTTGGAGCCCTTAAGAGGCTGT 1369  
1203 CTGGTGCATGCTGCAACTTTGGGGGAAACATGCTCTTTTGGAGCCCTTAAGAGGCTGT 1262  
1370 GAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1429  
1263 GAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1322  
1430 GGGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1489  
1323 AGCCCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1382  
1490 GCGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1549  
1383 GCGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1442  
1550 TGGGGTCTCCACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1609

1443 CGGGGCTCTCCAGGCTGATGCGGCTGGAGGCTTGGAGGCTCTCAATCAAGATGTCTCAAA 1502  
1610 CTGCTCCGGGAGGAGGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1669  
1503 CTGCTCTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1562  
1670 ACAGATGATATACCAATCATCTGATCAACCGATGCTGATGCTGATGCTGATGCTGATGCTGAT 1729  
1563 ACAGATGATATACCAATCATCTGATCAACCGATGCTGATGCTGATGCTGATGCTGATGCTGAT 1622  
1730 GCTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789  
1623 GTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1682  
1790 CACTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1849  
1683 CACCCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1742  
1850 CCTGAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909  
1743 CCTGAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1802  
1910 GCGGCACTGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1969  
1803 GCTTCACTAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1862  
1970 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2029  
1863 TCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1922  
2030 CTACCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2089  
1923 CTACCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1982  
2090 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2149  
1983 GCGAG 2042  
2150 TGACAGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2209  
2043 TCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2102  
2210 ACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2269  
2103 ACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2162  
2270 TGTGAGCCTGGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2329  
2163 CGTGAGCCTCTCCAGAGAGATCTTCTCAATATACACCC-----CAGCCTGACTCTTT 2216  
2330 GTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2389  
2217 GTGACGATTAACCATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2246

RESULT 7  
HSU43572 10127 bp DNA linear PRI 11-JUN-1996  
LOCUS HSU43572  
DEFINITION Human alpha-N-acetylglucosaminidase (NAGLU) gene, complete cds.  
ACCESSION U43572  
VERSION U43572.1 GI:1171228  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eubacteria; Primates; Carnivora; Homnidae; Homo.  
REFERENCE  
AUTHORS Zhao,H.G., Li,H.H., Bach,G., Schmidchen,A. and Neufeld,E.F.  
TITLE The molecular basis of Sanfilippo syndrome type B  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (12), 6101-6105 (1996)  
MEDLINE 96234097  
PUBMED 8650226



REFERENCE 2 (bases 1 to 10127)  
 AUTHORS Zhao, H.G., Li, H.H. and Neufeld, E.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-DEC-1995) Hong G. Zhao, Biological Chemistry, UCLA,  
 10833 Le Conte Ave., Los Angeles, CA 90095, USA  
 FEATURES  
 source  
 1. 10127  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17q21"  
 /cell\_type="fibroblast cell"  
 1085..3588  
 /gene="NAGLU"  
 join(1085..1799,2542..2689,3482..3628,3813..3898,  
 6091..6347,8167..9588)  
 /gene="NAGLU"  
 /note="5' end of mRNA is uncertain"  
 join(1417..1799,2542..2689,3482..3628,3813..3898,  
 6091..6347,8167..9377)  
 /gene="NAGLU"  
 /EC\_number="3.2.1.50"  
 /function="one of four enzymes involved in the degradation  
 of heparan sulfate; specifically removes the  
 alpha-N-acetylglucosamine residues"  
 /note="lysosomal enzyme; deficient in Sanfilippo B  
 syndrome"  
 /codon\_start=1  
 /product="alpha-N-acetylglucosaminidase"  
 /protein\_id="AAC50512.1"  
 /db\_xref="GI:1171229"  
 /translation="MEAVAVAAVGVLLIAGAGAGDEAREAAVRLVRLGPGP  
 AADPSYERALAAKRGDLSIGSGGAPRVRGSTVAAAGLRLRPGCHVA  
 WGSQRLRPPLPAVFGELTEATPNRRTYQVCTQSYSTWMDAREREDMALN  
 GNLALMSGQAIWQRYVALGLTQAEINEFTGPAPLAWGRMNLHWGPPEPS  
 HIKQLTLQRLVDQMSFGMTPEVLPAFAHVEAVTRFPQVNVMSQWCHFNCSYS  
 CEFLLADEDFIPIISLFLRLKEFGDHIYGA DTNEMQPSSESYLAATAY  
 YEAMPVDEAVLLOGLFQHPQPGPAQIRAVLGA VPRGLLVLLFASQVYV  
 RTASFGQPFIMQMLNFGNGLFGALFNAVNGPBARLFPNSTWGTGAPESISO  
 NRVVYSIMAEIGMRDVPDLAAVMTSPARRYGVSHDAAARLLISVYNGGA  
 CCHNRSPLVRPSLQNTSITVNRSDPEARLILTSAPLSLSPARILDLTRO  
 AVQELVSLYERASVLSKELASLRAGVLA YELLPLADVLSDRFLGSLTEQ  
 AAAAAVSEADFEYQNSRYQLTWPGENIIDLANKRLAGLVANVYPRRLFLEAL  
 VSVVAGIPIFOQHOPKXNVFOLEQAFVLSKQRYPSGPRGDTVDLAKKIFLKYYPGVVA  
 GSW"  
 1800..2541  
 /gene="NAGLU"  
 /number=1  
 /note="the sequence for this intron was partly determined  
 by the submitters and partly derived from GenBank  
 Accession Number M84472"  
 /number=2  
 3629..3812  
 /gene="NAGLU"  
 /number=3  
 3899..6090  
 /gene="NAGLU"  
 /note="the sequence for this intron was partly determined  
 by the submitters and partly derived from GenBank  
 Accession Number M84472"  
 /number=4  
 6348..8166  
 /gene="NAGLU"  
 /note="the sequence for this intron was partly determined  
 by the submitters and partly derived from GenBank  
 Accession Number M84472"  
 /number=5  
 BASE COUNT 2139 a 2846 c 2824 g 2318 t  
 ORIGIN

Query Match 55.4%; Score 1426.6; DB 9; Length 10127;  
 Best Local Similarity 99.7%; Pred. No. 5,2e-181;  
 Matches 1429; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 1112 CATATGTCAGTGTGATCTAGAGCTGTGTGCTCTCCAGAGCTGTGCTCTTCCAGACCA 1171  
 DB CTTGTCCACAGTGTGATCTAGAGCTGTGTGCTCTCTCCAGAGCTGTGCTCTTCCAGACCA 8215  
 1172 GCCGCACTTCTGCGGCGCCGCCAGATCAGAGCTGTGTGCTGTGAGCTGTGCTGTGCGG 1231  
 DB GCCCAGCTTCTGCGGCGCCGCCAGATCAGAGCTGTGTGCTGTGAGCTGTGCTGTGCGG 8275  
 1232 CCTCTGCTTCTGAGCTGTGTGCTGTGAGCAGAGCTGTGTATACCGGACTGTGCTCTT 1291  
 DB CTTCTGCTTCTGAGCTGTGTGCTGTGAGCAGAGCTGTGTATACCGGACTGTGCTCTT 8335  
 8276 CTTCTGCTTCTGAGCTGTGTGCTGTGAGCAGAGCTGTGTATACCGGACTGTGCTCTT 8335  
 1292 CCAGGCGCAGCCCTTCACTGTGTGATGTCACAACTTTGGGGGAAACCATGCTCTTT 1351  
 DB CCAGGCGCAGCCCTTCACTGTGTGATGTCACAACTTTGGGGGAAACCATGCTCTTT 8395  
 8336 CCAGGCGCAGCCCTTCACTGTGTGATGTCACAACTTTGGGGGAAACCATGCTCTTT 8395  
 1352 TGGAGCCCTTGAAGCTGTGAACCGAGGCCAGAGAGCTGCCGCTCTTCCCACTCCAC 1411  
 DB TGGAGCCCTTGAAGCTGTGAACCGAGGCCAGAGAGCTGCCGCTCTTCCCACTCCAC 8455  
 8396 TGGAGCCCTTGAAGCTGTGAACCGAGGCCAGAGAGCTGCCGCTCTTCCCACTCCAC 8455  
 1412 CATGTAGGACAGCGGCATGCGCCGCCAGGCAATCAGCAGAGAGTGTATTCTCT 1471  
 DB CATGTAGGACAGCGGCATGCGCCGCCAGGCAATCAGCAGAGAGTGTATTCTCT 8515  
 8456 CATGTAGGACAGCGGCATGCGCCGCCAGGCAATCAGCAGAGAGTGTATTCTCT 8515  
 1472 CATGGCTGAGCTGGCTGTGGGAAAGACCCAGTGTCCAGATTGGCAGCTGTGATCCAG 1531  
 DB CATGGCTGAGCTGGCTGTGGGAAAGACCCAGTGTCCAGATTGGCAGCTGTGATCCAG 8575  
 8516 CATGGCTGAGCTGGCTGTGGGAAAGACCCAGTGTCCAGATTGGCAGCTGTGATCCAG 8575  
 1532 CTTGGCCGCCCGGCGGATGATGGGCTCTCCAGCCGAGCAGCGGCGAGCTGTGAGCTACT 1591  
 DB CTTGGCCGCCCGGCGGATGATGGGCTCTCCAGCCGAGCAGCGGCGAGCTGTGAGCTACT 8635  
 8576 CTTGGCCGCCCGGCGGATGATGGGCTCTCCAGCCGAGCAGCGGCGAGCTGTGAGCTACT 8635  
 1592 GCTCCGAGAGTGTGTAACACTGTCTCCGGGAGGCTCTGAGGCGCCCAACTGTGACCCGCT 1651  
 DB GCTCCGAGAGTGTGTAACACTGTCTCCGGGAGGCTCTGAGGCGCCCAACTGTGACCCGCT 8695  
 8636 GCTCCGAGAGTGTGTAACACTGTCTCCGGGAGGCTCTGAGGCGCCCAACTGTGACCCGCT 8695  
 1652 GGTAGAGCGGCGGCTCTTCAAGATGAATACAGATCTGTGTAACAACGATCTGATGTGT 1711  
 DB GGTAGAGCGGCGGCTCTTCAAGATGAATACAGATCTGTGTAACAACGATCTGATGTGT 8755  
 8696 GGTAGAGCGGCGGCTCTTCAAGATGAATACAGATCTGTGTAACAACGATCTGATGTGT 8755  
 1712 TGAAGCTGTGCGGCTGTGCTCAATCTGCTCCCTCCGAGCAGCAGCCCGCTTCG 1771  
 DB TGAAGCTGTGCGGCTGTGCTCAATCTGCTCCCTCCGAGCAGCAGCCCGCTTCG 8815  
 8756 TGAAGCTGTGCGGCTGTGCTCAATCTGCTCCCTCCGAGCAGCAGCCCGCTTCG 8815  
 1772 CTACGACCTGTGACCTCACTCGGCGAGGACGTGACAGCTGTGATCTGATGA 1831  
 DB CTACGACCTGTGACCTCACTCGGCGAGGACGTGACAGCTGTGATCTGATGA 8875  
 8816 CTACGACCTGTGACCTCACTCGGCGAGGACGTGACAGCTGTGATCTGATGA 8875  
 1832 GGAAGCAAGAGCGCTCTACCTGAGCAGAGCTGTGCTCCCTGTGTGAGGCTGTGAGCGT 1891  
 DB GGAAGCAAGAGCGCTCTACCTGAGCAGAGCTGTGCTCCCTGTGTGAGGCTGTGAGCGT 8935  
 8876 GGAAGCAAGAGCGCTCTACCTGAGCAGAGCTGTGCTCCCTGTGTGAGGCTGTGAGCGT 8935  
 1892 CTTGCTTATGAGCTGTGCTGCGGCACTGACAGAGTGTGTGATGATCAACGCGCTTCTT 1951  
 DB CTTGCTTATGAGCTGTGCTGCGGCACTGACAGAGTGTGTGATGATCAACGCGCTTCTT 8995  
 8936 CTTGCTTATGAGCTGTGCTGCGGCACTGACAGAGTGTGTGATGATCAACGCGCTTCTT 8995  
 1952 GCTGGGAGCTGTGATAGAGGCGCGAGCAGGCGGCGAGTGTGAGGCGGAGCGGATTT 2011  
 DB GCTGGGAGCTGTGATAGAGGCGCGAGCAGGCGGCGAGTGTGAGGCGGAGCGGATTT 9055  
 8996 GCTGGGAGCTGTGATAGAGGCGCGAGCAGGCGGCGAGTGTGAGGCGGAGCGGATTT 9055  
 2012 CTACGACAGAGCAGCGCTTACGACTGACCTTGTGGGGCCAGAGGCAACATCTGGA 2071  
 DB CTACGACAGAGCAGCGCTTACGACTGACCTTGTGGGGCCAGAGGCAACATCTGGA 9115  
 9056 CTACGACAGAGCAGCGCTTACGACTGACCTTGTGGGGCCAGAGGCAACATCTGGA 9115  
 2072 CTATGCGAACAAGAGCTGTGCGGGGTTGTGGGCAATTAACACCTGTGTGCGGCT 2131  
 DB CTATGCGAACAAGAGCTGTGCGGGGTTGTGGGCAATTAACACCTGTGTGCGGCT 9175  
 9116 CTATGCGAACAAGAGCTGTGCGGGGTTGTGGGCAATTAACACCTGTGTGCGGCT 9175  
 2132 TTTCGTGAGGCGCTGTGTTGACAGTGTGCGCCAGGCGATCCCTTCCAGACGACAGTT 2191



```

Db      9176  TTTCTGAGAGGCGCTGATGACAGTGTGGCCAGGAGCCCTTCCAAACGACAGCAATT 9235
Qy      2192  TGACAAAATATGCTTCCAACTGAGACAGCCCTTCTGTTCTCAGACAGAGAGTACCCGAG 2251
Db      9236  TGACAAAATATGCTTCCAACTGAGACAGCCCTTCTGTTCTCAGACAGAGAGTACCCGAG 9235
Qy      2252  CCAGCCGCGAGAGACACTGTGTGACCTGAGCAAGAAATCTTCTCAATATTAACCCCG 2311
Db      9236  CCAGCCGCGAGAGACACTGTGTGACCTGAGCAAGAAATCTTCTCAATATTAACCCCG 9235
Qy      2312  CTGGGTGGCCGCTCTGTGTGATGATTCGACCAACAGCTGGGCTTGTCTTCGCTAATTC 2371
Db      9356  CTGGGTGGCCGCTCTGTGTGATGATTCGACCAACAGCTGGGCTTGTCTTCGCTAATTC 9415
Qy      2372  CAGGCGAGATTCAGAGGCGCCAGAGCTGACAGACATCAAGAGATTAACCCAGCCTGGAG 2431
Db      9416  CAGGCGAGATTCAGAGGCGCCAGAGCTGACAGACATCAAGAGATTAACCCAGCCTGGAG 9475
Qy      2432  CAGGCGCGACGCTGCTGTGTGTGGGTCTGACCTGGGGGGATTTGAGGGAAATGACCTGCC 2481
Db      9476  GAGGCGCGACGCTGCTGTGTGTGGGTCTGACCTGGGGGGATTTGAGGGAAATGACCTGCC 9535
Qy      2492  CTCACACACACACCAAGATGTGGATTAAGTACTGTTTCTTCCACTTAA 2544
Db      9536  CTCACACACACACCAAGATGTGGATTAAGTACTGTTTCTTCCACTTAA 9586

RESULT 8
ARI60673  ARI60673  10380 bp  DNA  linear  PAT 17-OCT-2001
LOCUS     ARI60673
DEFINITION ARI60673
ACCESSION ARI60673
VERSION   ARI60673.1  GI:16225098
KEYWORDS
ORGANISM  Unknown.
SOURCE    Unknown.
REFERENCE 1 (bases 1 to 10380)
AUTHORS   Hopwood, J. Joseph., Scott, H. Steele., Weber, B., Blanch, L. and
          Anson, D. Stewart.
          Synthetic mammalian .alpha.-n-acetylglucosaminidase and genetic
          sequences encoding same
          Patent: US 6255096-A 3 03-JUL-2001;
          Location/Qualifiers
            source          1..10380
                           /organism="Unknown"
BASE COUNT 2210 a 2952 c 2852 g 2366 t

Query Match      55.4%; Score 1426.6; DB 6; Length 10380;
Best Local Similarity 99.7%; Pred. No. 5.1e-181;
Matches 1429; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy      1412  CATGTAGGACACGAGCATGATGCCCGAGAGGCATACAGCAAGATGATATTCCT 1471
Db      8034  CATGTAGGACACGAGCATGATGCCCGAGAGGCATACAGCAAGATGATATTCCT 8093
Qy      1472  CATGTAGGACGAGGTGAGCAAAAGGATCCATGTGCATTTGGAGAGCTGGGTACAG 1531
Db      8094  CATGTAGGACGAGGTGAGCAAAAGGATCCATGTGCATTTGGAGAGCTGGGTACAG 8153
Qy      1532  CTTTGGCCGCGCGGATGAGGATCTCCACCCGAGCAGAGGAGAGAGCTGAGGCTACT 1591
Db      8154  CTTTGGCCGCGCGGATGAGGATCTCCACCCGAGCAGAGGAGAGAGCTGAGGCTACT 8213
Qy      1592  GCTCCGAGATGTGTACATCTCTCGGGAGAGGCTGACAGGGGCAATCTGATGCCGT 1651
Db      8214  GCTCCGAGATGTGTACATCTCTCGGGAGAGGCTGACAGGGGCAATCTGATGCCGT 8273
Qy      1652  GGTCAAGCGGCGGTCTCAAGATGAATACAGCATCTGTACACCATCTGATGTGT 1711
Db      8274  GGTCAAGCGGCGGTCTCAAGATGAATACAGCATCTGTACACCATCTGATGTGT 8333
Qy      1712  TGAAGCTGAGCGGCTGTGCTCAATCTGCTCTCTCTGCTGCTGCTGCTGCTGCTG 1771
Db      8334  TGAAGCTGAGCGGCTGTGCTCAATCTGCTCTCTCTGCTGCTGCTGCTGCTGCTG 8393
Qy      1772  CTACGACCTGTGAGACTGACTCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1831
Db      8394  CTACGACCTGTGAGACTGACTCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8453
Qy      1832  GGAGGCAAGAGAGGCTACTCCTAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891
Db      8454  GGAGGCAAGAGAGGCTACTCCTAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8513
Qy      1892  CTTGGCTTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1951
Db      8514  CTTGGCTTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8573
Qy      1952  GCTGGGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2011
Db      8574  GCTGGGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8633
Qy      2012  CTAGGACAGAAACAGCGCTACCAAGCTGACCTTGTGGGGGCGAAGAGCAATCTTGA 2071
Db      8634  CTAGGACAGAAACAGCGCTACCAAGCTGACCTTGTGGGGGCGAAGAGCAATCTTGA 8693
Qy      2072  CTATGCGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131
Db      8694  CTATGCGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8753
Qy      2132  TTTCTGAGAGCGCTGTGACAGTGTGCGCAAGAGAGATCCCTTTCACACAGCAGTT 2191
Db      8754  TTTCTGAGAGCGCTGTGACAGTGTGCGCAAGAGAGATCCCTTTCACACAGCAGTT 8813
Qy      2192  TGACAAAAATATGCTTCCAACTGAGACAGCCCTTCTGTTCTCAGACAGAGTACCCG 2251
Db      8814  TGACAAAAATATGCTTCCAACTGAGACAGCCCTTCTGTTCTCAGACAGAGTACCCG 8873
Qy      2252  CCAGCCGCGAGAGACACTGTGTGACCTGAGCAAGAAATCTTCTCAATATTAACCCCG 2311
Db      8874  CCAGCCGCGAGAGACACTGTGTGACCTGAGCAAGAAATCTTCTCAATATTAACCCCG 8933
Qy      2312  CTGGGTGGCCGCTCTGTGTGATGATTCGACCAACAGCTGGGCTTGTCTTCGCTAATTC 2371
Db      8934  CTGGGTGGCCGCTCTGTGTGATGATTCGACCAACAGCTGGGCTTGTCTTCGCTAATTC 8993
Qy      2372  CAGGCGAGATTCAGAGGCGCCAGAGCTGACAGACATCAAGAGATTAACCCAGCCTGGAG 2431
Db      8994  CAGGCGAGATTCAGAGGCGCCAGAGCTGACAGACATCAAGAGATTAACCCAGCCTGGAG 9053
Qy      2432  GAGGCGCGACGCTGCTGTGTGTGGGTCTGACCTGGGGGGATTTGAGAGAAATGACCTGCC 2491
Db      9054  GAGGCGCGACGCTGCTGTGTGTGGGTCTGACCTGGGGGGATTTGAGAGAAATGACCTGCC 9113

```

QY 2492 CTCACACACACCAAGTGGATTAAAGTACTGTTTCTTCCACTTAA 2544  
 Db 9114 CTCACACACACCAAGTGGATTAAAGTACTGTTTCTTCCACTTAA 9166

RESULT 9  
 HSU34879 46610 bp DNA linear PRI 14-JAN-1999  
 LOCUS Human 17-beta-hydroxysteroid dehydrogenase (BDH17B2) gene, complete  
 DEFINITION cds.  
 ACCESSION U34879  
 VERSION U34879.1 GI:975694  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 46610)  
 Zhao, Z., Yazdani, A., Shen, Y., Sun, Z., Bailey, J., Caskey, C.T. and  
 Lee, C.C.  
 REFERENCE Molecular dissection of a cosmid from a gene-rich region in 17q21  
 and characterization of a candidate gene for  
 alpha-N-acetylglicosaminidase with two cDNA isoforms  
 Mamm. Genome 7 (9), 666-690 (1996)  
 PUBMED 8703123  
 JOURNAL  
 MEDLINE  
 REFERENCE 2 (bases 1 to 46610)  
 AUTHORS Shen, Y. and Gibbs, R.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-AUG-1995) Ying Shen and Richard A. Gibbs, Molecular &  
 Human Genetics, Baylor College of Medicine, One Baylor Plaza,  
 Houston, TX 77030, USA

FEATURES  
 source  
 1. 46610  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17q21.1"  
 /clone="140C10"  
 /clone\_1b="Los Alamos Cosmid Library: LANCO17"  
 complement(356..528)  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 complement(529..806)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(1082..1412)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(1414..1749)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 2374..2717  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 complement(2868..3078)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 3232..3631  
 /rpt\_family="MER9"  
 /rpt\_type=direct  
 complement(3633..3882)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 4239..4535  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 4668..4922  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 complement(5694..5924)  
 /rpt\_family="Alu"

repeat\_unit  
 complement(5958..6237)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(7197..7485)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(7489..7796)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(7804..8085)  
 /rpt\_family="MSART1"  
 /rpt\_type=inverted  
 8900..9066  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 complement(9119..9309)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(9310..9597)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(9599..9742)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 12373..12557  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 12635..34401  
 /note="region overlaps with GenBank Accession Number  
 M84472"  
 14510..14805  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 14839..15086  
 /rpt\_family="D1"  
 15094..15389  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 complement(16583..16749)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(16750..17039)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(17179..17466)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(19249..19418)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(19419..19713)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 19918..20162  
 /rpt\_family="MER21"  
 /rpt\_type=direct  
 complement(20167..20504)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 20748..20899  
 /rpt\_family="MER40"  
 /rpt\_type=direct  
 23522..23755  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 complement(23988..24293)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(25635..35976)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 25854..26145

Query Match	Best Local Similarity	Matches 1429; Conservative	Score 1426.6; DB 9; Length 46610; Pred. No. 3.6e-181; Mismatches 4; Indels 0; Gaps 0;
DB 1112	CATGACTCAGTGTACTGAGGCTGTGTGCTGTCTCCAGGCTGGCTCTTCACAGCA	1172	1172
DB 17694	CCTGTCCACAGTGGATCTGAGGCTGTGTGCTGTCTCCAGGCTGGCTCTTCACAGCA	17753	17753
DB 1172	GCCGCAAGTTCTGGGGGCCGCCCAATCAGGAGCTGTGCTGGGAGCTGTGCTGGGCGC	1231	1231
DB 17754	GCCGCAAGTTCTGGGGGCCGCCCAATCAGGAGCTGTGCTGGGAGCTGTGCTGGGCGC	17813	17813

QY	1232	CCTCCTGTTTGTGAAGCTGTTTGTGTGAAGCAGACCTGTGTATATCCCGACCTGCCTTCCT	1291
Db	17814	CCTCCTGTTTGTGAAGCTGTTTGTGTGAAGCAGACCTGTGTATATCCCGACCTGCCTTCCT	17877
QY	1292	CCAGGGCAGGCGCTTCACTGTGTGCATGTCTGCACAACTTTGTGGGGAAACCATGATCTTTT	1351
Db	17874	CCAGGGCAGGCGCTTCACTGTGTGCATGTCTGCACAACTTTGTGGGGAAACCATGATCTTTT	17933
QY	1352	TGAGCCCTTAGAGGCTGTGAACGGAGGCCAGAAAGCTGCCGCTCTTCCCACTTCAC	1411
Db	17934	TGAGCCCTTAGAGGCTGTGAACGGAGGCCAGAAAGCTGCCGCTCTTCCCACTTCAC	17993
QY	1412	CATGGTGAAGCAGCGGGCATTGGCCCCGAGGGGCATTCAGCAGAAAGAGTGTCTATTCCT	1471
Db	17994	CATGGTGAAGCAGCGGGCATTGGCCCCGAGGGGCATTCAGCAGAAAGAGTGTCTATTCCT	1805
QY	1472	CATGGCTGAGCTGGGCTGGCGAAAGAACCCAGTGCAGATTTTGGACAGCCTGGGGTGAACAG	1531
Db	18054	CATGGCTGAGCTGGGCTGGCGAAAGAACCCAGTGCAGATTTTGGACAGCCTGGGGTGAACAG	18111
QY	1532	CTTTGGCCGCCCGGGGGGTATGGGGTCTCCCAACCCGAGACGAGGGGGCAGCGTGGAGGCTACT	1591
Db	18114	CTTTGGCCGCCCGGGGGGTATGGGGTCTCCCAACCCGAGACGAGGGGGCAGCGTGGAGGCTACT	18177
QY	1592	GCTCCGAGAGTGTGTACAACCTGCTCCGGGGAGGCGCTGAGGGGCGCAAACTGTAGCCGCT	1651
Db	18174	GCTCCGAGAGTGTGTACAACCTGCTCCGGGGAGGCGCTGAGGGGCGCAAACTGTAGCCGCT	18233
QY	1652	GGTCAAGCGCGGCGCTCCTTACAGATGAATACAGACTTGTGTACACCGACTCATATGTCTT	1711
Db	18234	GGTCAAGCGCGGCGCTCCTTACAGATGAATACAGACTTGTGTACACCGACTCATATGTCTT	18293
QY	1712	TGAGGCGCTGGGGGTGTGTGCTCAACTGTGCTCCCTCCCTGGCCAGAGCCCGCCTTCG	1771
Db	18294	TGAGGCGCTGGGGGTGTGTGCTCAACTGTGCTCCCTCCCTGGCCAGAGCCCGCCTTCG	18355
QY	1772	CTACGACCTGTGGAACCTCACTTCGGCAGGCAGTGCAGAGACTGTCACTTGTACTATGA	1831
Db	18354	CTACGACCTGTGGAACCTCACTTCGGCAGGCAGTGCAGAGACTGTCACTTGTACTATGA	18413
QY	1832	GGAAGGCAAGAAAGCGCCTACCTTGAGCAAGAGACTGGGCTCCCTGTAGAGGGCTGGAAGCGT	1891
Db	18414	GGAAGGCAAGAAAGCGCCTACCTTGAGCAAGAGACTGGGCTCCCTGTGTAGAGGGCTGGAAGCGT	18477
QY	1892	CCTGGCCTATAGAGCTGTGCTGGCACTGTGAAGAGGTGTCTGGTACTGTACAGCCGCTTCTT	1951
Db	18474	CCTGGCCTATAGAGCTGTGCTGGCACTGTGAAGAGGTGTCTGGTACTGTGTACAGCCGCTTCTT	18533
QY	1952	GCTGGGCAAGCTGGCTTAGAGCAAGGCCGAGAGTACGCGGAGTCACTGAAGCCGAGGCCATTT	2011
Db	18534	GCTGGGCAAGCTGGCTTAGAGCAAGGCCGAGAGTACGCGGAGTCACTGAAGGCCGAGGCCATTT	18593
QY	2012	CTACGACAGAGCAAGCGGCTACAGCTGTGCTGGGGGCGAGAAAGCAACTCTGTGA	2071
Db	18594	CTACGACAGAGCAAGCGGCTACAGCTGTGCTGGGGGCGAGAAAGCAACTCTGTGA	18655
QY	2072	CTATGCGAACAGAGCTGGCGGGGTTGTGTGSCAACTACTACACCCCTGCTGGCGCT	2131
Db	18654	CTATGCGAACAGAGCTGGCGGGGTTGTGTGSCAACTACTACACCCCTGCTGGCGCT	18713
QY	2132	TTTTCTGTGAAGCGCTGGTTTGAAGAGTGTGGCCAGAGGCATCCCTTCCACAGACCAAGTT	2191
Db	18714	TTTTCTGTGAAGCGCTGGTTTGAAGAGTGTGGCCAGAGGCATCCCTTCCACAGACCAAGTT	18777
QY	2192	TGACAAAATGTCTTCCAACTGAGCAAGGCTTTCGTTCTTACAGAACAGAGGTATCCCAAG	2251
Db	18774	TGACAAAATGTCTTCCAACTGAGCAAGGCTTTCGTTCTTACAGAACAGAGGTATCCCAAG	18833
QY	2252	CCAGCGCCGAGAGACACTGTGCACTTCGGCAGAAAGATCTTCCATAATATTAACCCGG	2311
Db	18834	CCAGCGCCGAGAGACACTGTGCACTTCGGCAGAAAGATCTTCCATAATATTAACCCGG	18893

QY 2312 CTGGGTCGGCGGCTCTGTGGATGATGATTGGCCACACTGGGCTTGTGTTTCCGCTAATTC 2371  
 DB 18894 CTGGGTCGGCGGCTCTGTGGATGATGATTGGCCACACTGGGCTTGTGTTTCCGCTAATTC 18953  
 QY 2372 CAGGGCAGATTCGAGGGGCGGAGAGCTGAGACATACAGAGGATTAACCCAGGCGCTGGAG 2431  
 DB 18954 CAGGGCAGATTCGAGGGGCGGAGAGCTGAGACATACAGAGGATTAACCCAGGCGCTGGAG 19013  
 QY 2432 GAGGCCCCCAGCGGCTGTGTGGGAGCTGACCTGGGGGAGATTGAGGAGGAATGACCTGCC 2491  
 DB 19014 GAGGCCCCCAGCGGCTGTGTGGGAGCTGACCTGGGGGAGATTGAGGAGGAATGACCTGCC 19073  
 QY 2492 CTCACACACACACACCAAGATGTGGATTAAGTACTGTTTCTTCCACTTAA 2544  
 DB 19074 CTCACACACACACCAAGATGTGGATTAAGTACTGTTTCTTCCACTTAA 19126  
 RESULT 10  
 AC067852/c 179146 bp DNA linear PRI 27-FEB-2003  
 LOCUS AC067852 Homo sapiens chromosome 17, clone RP11-400F19, complete sequence.  
 DEFINITION AC067852  
 ACCESSION AC067852 GI:28570403  
 VERSION AC067852.23  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
 Boguslavskiy, L., Bouckgalter, B., Brown, A., Burkett, G.,  
 Campopoli, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, D.,  
 Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
 Howland, J., C. Iliev, I., Johnson, R., Jones, C., Kann, L., Kartas, A.,  
 Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lemock, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., Mhova, T., McKernan, K., McPheeters, R.,  
 Meidirm, J., Meneus, L., Mhova, T., Miranda, C., Mlenga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 179146)  
 Biren, B., Nusbaum, C., Lander, E., Abouellell, A., Allen, N.,  
 Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,  
 Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y.,  
 Collymore, A., Cook, A., Cooke, P., Corum, B., Dearrellano, K.,  
 Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S.,  
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,  
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
 Meidirm, J., Meneus, L., Mhova, T., Mlenga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Topham, K., Travers, M.,  
 Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (31-JAN-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 179146)  
 Biren, B., Nusbaum, C., Lander, E., Abouellell, A., Allen, N.,  
 Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,  
 Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y.,  
 Collymore, A., Cook, A., Cooke, P., Corum, B., Dearrellano, K.,  
 Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S.,  
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,  
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
 Meidirm, J., Meneus, L., Mhova, T., Mlenga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Topham, K., Travers, M.,  
 Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (27-FEB-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L9335  
 Center clone name: 400\_F\_19  
 -----  
 FEATURES  
 source 1..179146  
 /organism="Homo sapiens"  
 /mol\_type="Genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17"  
 /clone="RP11-400F19"  
 /clone\_id="RP11-400F19 Human Male BAC"  
 138..440  
 /rpt\_family="AluYo"  
 /rpt\_family="GC\_rich"  
 962..1001  
 /rpt\_family="GC\_rich"  
 1058..1081  
 /rpt\_family="GC\_rich"  
 1327..1382  
 /rpt\_family="CT-rich"  
 2432..2469  
 /rpt\_family="GC\_rich"  
 2902..2934  
 /rpt\_family="AGGGGAGn"  
 /complement(3214..3365)  
 /rpt\_family="MIR3"  
 /complement(5216..5317)  
 /rpt\_family="MIR"  
 5700..5721  
 /rpt\_family="GC\_rich"  
 /complement(6114..6215)









gene 7730..13913  
       /gene="EDH17B1"  
 repeat\_unit 7730..13913  
       /gene="EDH17B1"  
       /note="G00-119-860"  
 misc\_difference 7802..7805  
       /gene="EDH17B1"  
       /note="present in clone lambda17-HSD5; absent in clone lambda17-HSD3"  
 misc\_difference 7809..7815  
       /gene="EDH17B1"  
       /note="present in clone lambda17-HSD5; absent in clone lambda17-HSD3"  
 prim\_transcript 7816..11014  
       /gene="EDH17B1"  
       /note="G00-119-860"  
 misc\_difference 8095  
       /gene="EDH17B1"  
       /note="G in clone lambda17-HSD5; A in clone lambda17-HSD3;  
       G00-119-860"  
 misc\_difference 8178  
       /gene="EDH17B1"  
       /note="C in clone lambda17-HSD5; T in clone lambda17-HSD3;  
       G00-119-860"  
 repeat\_unit 10880..11014  
       /gene="EDH17B1"  
       /note="left arm"  
       /rpc\_family="Alu"  
       /rpc\_type="direct  
       complement(11393..11650)  
 repeat\_unit  
       /rpc\_family="Alu"  
       /rpc\_type="inverted  
       1267  
 misc\_difference  
       /gene="EDH17B1"  
       /note="C in clone lambda17-HSD5; T in clone lambda17-HSD3"  
 repeat\_unit 13208..13930  
       /gene="EDH17B1"  
       /rpc\_family="Alu"  
       /rpc\_type="direct  
       13520..13794  
 repeat\_unit  
       /gene="EDH17B1"  
       /rpc\_family="Alu"  
       /rpc\_type="direct  
       13811..13913  
 repeat\_unit  
       /gene="EDH17B1"  
       /note="truncated"  
       /rpc\_family="Alu"  
       /rpc\_type="direct  
       13914..20156  
 gene  
       /gene="EDH17B2"  
       13914..20156  
 repeat\_unit  
       /gene="EDH17B2"  
       /note="G00-127-970"  
       complement(13914..14049)  
 repeat\_unit  
       /rpc\_family="Alu"  
       /rpc\_type="inverted  
       join(13999..15062,15156..15323,15471..15650,15826..15919,  
       16436..16613,16700..17244)  
       /gene="EDH17B2"  
       /note="G00-127-970"  
 prim\_transcript 13999..17244  
       /gene="EDH17B2"  
       /note="G00-127-970"  
 exon 13999..15062  
       /gene="EDH17B2"  
       /note="G00-127-970"  
 CAAT\_signal  
       complement(14865..14868)  
       /note="G00-127-970"  
 protein\_bind 14897..14898  
       /gene="EDH17B2"  
       /note="G00-127-970"  
       /bound\_moiety="transcription factor AP-2"

GC\_signal 14902..14915  
       /gene="EDH17B2"  
       /note="G00-127-970"  
 TATA\_signal 14925..14931  
       /gene="EDH17B2"  
       /note="G00-127-970"  
 CDS join(14966..15062,15156..15323,15471..15650,15826..15919,  
       16436..16613,16700..16969)  
       /gene="EDH17B2"  
       /EC\_number="1.1.1.62"  
       /standard\_name="estradiol 17-beta-dehydrogenase"  
       /codon\_start=1  
       /product="17-beta-hydroxysteroid dehydrogenase"  
       /protein\_id="AAB16942.1"  
       /db\_xref="GI:177127"  
       /translation="MARTVLTGSGSIGLHVALRASPQSFVYATLRLTKTG  
       RLMEARALACPGLSELTLDVDRDSKGVAAARVYVLTGAVGLGPLEA  
       LGEDAVASVLDVNVGTRMLQAFIDPKRRSGSRVLTGAVGLGPLEA  
       PALSELCSLAVLLPFGVHSLIECGVHTLPMKRYGSPREVTDRDITFHFYQ  
       YLAHKKQVREAAQNPSEVAVFTLALAPKFTLFTTTERFLPLRLRLDDPSGSNT  
       VTAMREVFQDVPAFAEAAGAAGAGGAGAEAGSAVGDPELDPPAPQ"  
       15063..15155  
       /gene="EDH17B2"  
       /note="G00-127-970"  
       /number=1  
 intron 15156..15323  
       /gene="EDH17B2"  
       /note="G00-127-970"  
       /number=2  
 exon 15324..15470  
       /gene="EDH17B2"  
       /note="G00-127-970"  
 intron  
       /note="G00-127-970"

Query Match 54.9%; Score 1414.6; DB 9; Length 21764;  
 Best Local Similarity 99.7%; Pred. No. 1.7e-179;  
 Matches 1428; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY	1112	CATGACTGAGTGAATGACTGAGCTGTGTGCTGCTCAAGGCTGCTTCCAGACCA	1171
DB	5056	CCTGTCCACAGTGAATGACTGAGCTGTGTGCTGCTCAAGGCTGCTTCCAGACCA	5115
QY	1172	GCAGCAGTTCTGAGGAGCCGCCAGATCAAGGCTGTGAGAGTGTGCCCCGTGGCCG	1231
DB	5116	GCAGCAGTTCTGAGGAGCCGCCAGATCAAGGCTGTGAGAGTGTGCCCCGTGGCCG	5175
QY	1232	CCTCTGCTTCTGACCTGTGTGAGGACCAAGCTGTGTATACCCGACCTGCTCCTT	1291
DB	5176	CCTCTGCTTCTGACCTGTGTGAGGACCAAGCTGTGTATACCCGACCTGCTCCTT	5235
QY	1292	CCAGGACCAAGCTTCACTGTGTGATGCTGCACAACTTTGGGGGAAACCATGCTTTT	1351
DB	5236	CCAGGACCAAGCTTCACTGTGTGATGCTGCACAACTTTGGGGGAAACCATGCTTTT	5295
QY	1352	TGAGAGCTTGAAGGCTGTGAACGAGAGCCCAAGCTGCCCTCTTCCCACTCCAC	1411
DB	5296	TGAGAGCTTGAAGGCTGTGAACGAGAGCCCAAGCTGCCCTCTTCCCACTCCAC	5355
QY	1412	CATGTAGGACAGGGGATGAGCCCGAGAGGACATCAGCCAGAACGATGTATTTCCT	1471
DB	5356	CATGTAGGACAGGGGATGAGCCCGAGAGGACATCAGCCAGAACGATGTATTTCCT	5415
QY	1472	CATGTAGGCTGGGGCTGCGGAAAGAACCACTGCGCAATTTGGAGCTTGAGTACAG	1531
DB	5416	CATGTAGGCTGGGGCTGCGGAAAGAACCACTGCGCAATTTGGAGCTTGAGTACAG	5475
QY	1532	CTTTGCGGCGCGGAGTATGGGGTCTCCACCGGACGAGAGGGGAGAGCTGAGGCTACT	1591
DB	5476	CTTTGCGGCGCGGAGTATGGGGTCTCCACCGGAGAGAGGGGAGAGCTGAGGCTACT	5535
QY	1592	GCTCCGAGTGTGTACCACTGTCTCCGGAGAGGCTTGACAGAGGCAATCGTATCCGCT	1651
DB	5536	GCTCCGAGTGTGTACCACTGTCTCCGGAGAGGCTTGACAGAGGCAATCGTATCCGCT	5595

OY		1692	GGTCAAGGGGCGCCGTCTCCTTACAGATGAATAACAGAAGCATCTGGTAACAACGCATCTGAATGTT	1711
Dd		5896	GCTAAGGCGGCCGTCCTCCCTCACATBAATAAATCCAAGCATCTGGTNCAACCACATCTGAATGTT	5655
OY		1712	TGAAGCCTGCGCGCTGCTGCTCACACTCTTGCTCCCTCCCTGGCCACAGCCCCGCTTCG	177U
Dd		5656	TGAGGCCCTGCGCGCTGCTGCTCACACTCTTGCTCCCTCCCTGGCCACAGCCCCGCTTCG	5715
OY		1772	CTAGAACCTGCTGAGACCTCACTCCGGCAGAGCTGAGAGAGCTGGTCAAGCTTGTACTAGA	1831J
Dd		5716	CTACGACCTGCTGAGACCTCACTCCGGCAGAGCTGAGAGAGCTGGTCAAGCTTGTACTAGA	5775
OY		1832	GGAGGCAGAAGCGCGCTACCTTAGCANAGAGCTGGCCTCCCTGTGTGAAGGCTGAGAGCGT	1891
Dd		5776	GGAGGCAGAAGCGCGCTACCTTAGCANAGAGCTGGCCTCCCTGTGTGAAGGCTGAGAGCGT	5835
OY		1892	CCTGGCTTATGAGAGCTGCTGCCGACCTGSAAGAGTGCTGGCTAGTGAAGCCGCTTCTT	1951
Dd		5896	CCTGGCTTATGAGAGCTGCTGCCGACCTGSAAGAGTGCTGGCTAGTGAAGCCGCTTCTT	5895
OY		1952	GCTGGGCGAGCTGGCTAGAGAGAGCCCAGACAGCGGCACTGACGTAGAGGCGGAGGCGATT	2011H
Dd		5896	GCTGGGCGAGCTGGCTAGAGAGAGCCCAGACAGCGGCACTGACGTAGAGGCGGAGGCGATT	5955
OY		2012	CTACGAGCAGAACAAGCGCGCTACAGACTGACCTTGTGGGGGCCAGAGGCCAATCTTGSA	2071U
Dd		5955	CTACGAGCAGAACAAGCGCGCTACAGACTGACCTTGTGGGGGCCAGAGGCCAATCTTGSA	6014
OY		2072	CTATGGCAAACAAGAGAGCTGGGGGGTTTTGGTGGCCAACTATACACCCCTCGTGGCGGCT	2131J
Dd		6015	CTATGGCAAACAAGAGAGCTGGGGGGTTTTGGTGGCCAACTATACACCCCTCGTGGCGGCT	6074
OY		2132	TTTCCTGGAGGCGCTGATTGACAGTGTGGCCCAAGGSCATCCCTTTCACAACAGACAGTT	2191
Dd		6075	TTTCCTGGAGGCGCTGATTGACAGTGTGGCCCAAGGSCATCCCTTTCACAACAGACAGTT	6134
OY		2192	TGACAAAAATGCTTTCCGACTGGAGACAGGCTTCGTTCTCAGACAGCAGAGTACCCAG	2251
Dd		6135	TGACAAAAATGCTTTCCGACTGGAGACAGGCTTCGTTCTCAGACAGCAGAGTACCCAG	6194
OY		2252	CCAGCCGCGAGAGAGACACTGTGTGACCTGGCCAAAGAGATCTTCCCAATATTACCCCG	2311H
Dd		6195	CCAGCCGCGAGAGAGACACTGTGTGACCTGGCCAAAGAGATCTTCCCAATATTACCCCG	6255
OY		2312	CTGGGTGGCCCGGCTTTGGTGTATGATTTGCGCACCACTGGGCTTGTCTTCOGCTAATTC	2371U
Dd		6255	CTGGGTGGCCCGGCTTTGGTGTATGATTTGCGCACCACTGGGCTTGTCTTCOGCTAATTC	6314
OY		2372	CAGGGCAGATTCCAGGGGCCAGAGCTGGACAGACATCAAGAGTAAACCAAGCCTGGGAG	2431
Dd		6315	CAGGGCAGATTCCAGGGGCCAGAGCTGGACAGACATCAAGAGTAAACCAAGCCTGGGAG	6374
OY		2432	GAGGCCCAAGCGGCTGCTGTGTGGGCTTGACCTGGGGGGATTGAGGGAAAATGACTTGCC	2491
Dd		6375	GAGGCCCAAGCGGCTGCTGTGTGGGCTTGACCTGGGGGGATTGAGGGAAAATGACTTGCC	6434
OY		2492	CTCCACCAACCAACCCCAAGGTGGGATTAAAGTACGTTTCTTTCOCATTAA	2544
Dd		6435	CTCCACCAACCAACCCCAAGGTGGGATTAAAGTACGTTTCTTTCOCATTAA	6487

RESULT\_13  
AF331668  
LOCUS AF331668  
DEFINITION Dromaius novaehollandiae lysosomal alpha-N-acetyl glucosaminidase  
VERSION AF331668  
KEYWORDS AF331668.1 GI:14861377  
SOURCE Dromaius novaehollandiae (emu)  
ORGANISM Dromaius novaehollandiae (emu); Eukaryota; Metazoa; Chordata; Dromaiidae; Archaeosauria; Aves; Palaeognathae; Casuariiformes; Dromalidae;

REFERENCE	Dromaius.
AUTHORS	1 (bases 1 to 2949)
TITLE	Aronovich, E.L., Johnston, J.M., Wang, P., Giger, U. and Whitey, C.B.
JOURNAL	Molecular basis of mucopolysaccharidoses type IIIB in emu ( <i>Dromaius novaehollandiae</i> ): an avian model of Sanfilippo syndrome type B
MEDLINE	Genomics 74 (3), 299-305 (2001)
PIUMED	21309063
REFERENCE	11414757
AUTHORS	2 (bases 1 to 2949)
TITLE	Aronovich, E.L. and Whitey, C.B.
REFERENCE	Direct Submission
JOURNAL	Submitted (21-DEC-2000) Peds & Inst. Human Genetics, University of Minnesota, 420 Delaware St. SE, Minneapolis, MN 55455, USA
FEATURES	Location/Qualifiers
source	1..2949
	/organism="Dromaius novaehollandiae"
	/mol_type="mRNA"
	/db_xref="taxon:8790"
	13..2274
	/note="NAGLU"
	/codon_start=1
	/product="lysosomal alpha-N-acetyl glucosaminidase"
	/protein_id="AAK73654.1"
	/db_xref="GI:14861378"
	/translation="MAAPRVRLTALAAMAAARAAPLPVYRAGAEADARQEAVALA RLLGPRAAVALSDGSLAAGIDTRVHSPGAAYAVVAVAGSGVAALAGHYR DLCSHLSMGSRGRLPLDPLPRVPAEIRAPGFRYQVCTOSYEFAMDMARWE EIDMGLSGINILALAFAGOEAVMORVLSIGLNSIDEVETGAPLANRMGNLHG AGPLPRAMHKOLVVOYRVLEPMSSLCMTIYLPFAGHVPGVRAIPRVATPLGK SHPECTYSCITLDPEDPMVOYITLPLKELIKRFGRDHTLSADTNEMNTSSDPA LSRVSAVFRSMGTADPKAVLMQGMILFOHPDPMQAOVRALHGFPLGERTVDLE AESNPVQWTESFYGPPFIWCMENLFGSNGLPETAVINHGPAARFPNSTVGTG LVPEGLIQNDMVFYGLNMLQMRQEPDLPSVAIVARRYGAENPAALASAWOLLRS YNTGCVQVNNHRSPLVLRPSLMDTEVYNKSDVYEAMRLLSGAEIGSSPTGYDI ADVTROAQGLVSEYLYSIRQAFOSRSLPELLTAGGLVYDLLELDGLSSHLFL GRMESLRAVTSRBEAOYEINARNQVTIMGNKTLIDYANKRGLGVLIDYGVNMS LPSVAIVESLNSGSPFHQDFQNAVPQVNGFIYNNKRYPIAPVGTLEISKILFKK YPSAMRRGRAGPA
BASE COUNT	425 a 1025 c 1038 g 461 t
ORIGIN	
Query Match	43.3%; Score 114.2; DB 5; Length 2949;
Best Local Similarity	69.6%; Pred. No. 2.8e-119;
Matches 1541; Conservative	0; Mismatches 668; Indels 6; Gaps 2;
QY	98 GACCATGAGGCGGTGGCGGTGGCGGCGGCGGCGGCGGCTCTTCTCTGCGCGGCGCGG 157
DB	33 GACGCGGTGGCTGCTCCATGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 92
QY	158 GGGCGCGGCGAGCGACAGAGCGCGGCGGAGAGCGGCGGCGGCGGCGGCGGCGGCT 217
DB	93 GCGGCGCGGCGCGGAGGAGCGGCGGCGGAGAGGCGGCGGCGGCGGCGGCGGCGCT 152
QY	218 GCTGGGCGAGGCGCGCGGCGGCGGCTTCTCGTTCGATGAGAGCGCGGCTTGAGTCCCA 277
DB	153 GCTGGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 212
QY	278 GCCGCGCTTGACACCTACAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 337
DB	213 GCTGAGACCTACCGGCGGAGCTGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 272
QY	338 CTCACGCGGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 397
DB	273 CTCAGCGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 332
QY	398 CCACGTGGCGTGTTCGGCTCTACAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 457
DB	333 CCACCTCTCCCTGTGTTCGGGAGGAGCTTCGCTCGCGGAGCCCGGCGGCGGCGG 392
QY	458 GGACGTGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 517
DB	393 CGAATCGCGCGCACCGCGCGCGGCGGCGGCTTCGCTACTACAGAAAGCTCTCACCCAGG 452

QY	5148	CTACTCCCTTCGCTGGGTGGGACTCGGGCCCGCTGAGGCGAGATGACATGCGATGGCGCT	577
Db	4453	CTACTCGTTCGCGTGTGGGACTCGGGCGCGCTGGGAGAGAGATGACTGATGGACCTCT	512
QY	578	GATGGCATCAACTTGAGCACTGGCTTGAGACGGCCAGAGGCCATCTGGACGGGTGTGA	637
Db	513	CAGCGGCATCAACTTGAGCACTGGCGCTTGAGAGGCCATCTGGACGGGTGTGA	572
QY	638	CGTGGCTTGGGGCTGAGCCAGGCAAGATGAGTGTCTTATCTGCTCTGCTCTCT	697
Db	573	CCTTCTCTTGGGCTTCAACCAATGCGAGATGACGAGTATTTTCAACGGGGCCGCGTTCCT	632
QY	698	GGCCTGGGGCGAATGGGCACCTTGACACTTGAGATGGCCCTTCGCCCTCTGGCA	757
Db	633	GGCCTGGGACCGATGGGAAACCTTCACGCGCTGGGACGGGGCGCTGCGCGGCTGGCA	692
QY	758	CATCAAGCAGCTTTAACCTGACAGCAGGGTCTTGACCAAGATGCGCTCTTGCGCATGAC	817
Db	693	CTTCAAGCAGCTTTAAGTGCAGATGACAGGGTCTTGAGAGAGATGCGCTCCCTGGGAGTAT	752
QY	818	CCCAAGTGCCTGCAATTCGCGGGGAGATGTTCCCGAGCGTGCACAGGAGTGTCCCTCA	877
Db	753	CACGGTGCCTGGCGCTTGACAGGCCACGTGCCCCAGGCGTCTTAGAGACTTTCACAG	812
QY	878	GGTCAATGTCAGAAAGATGGGCAAGTTGGGCGCACTTAACTGTCTACTCTGCTCTT	937
Db	813	TGTGAATGCACTGCGCTCGGGGGCTGAGGCCACTTTGACTGCAGATCTTTGTAACTGA	872
QY	938	CCCTTCGGCTCGGAAAGACCCCATATTCCTCCATATCCGAGGACCTCTTCTGCGAGCT	997
Db	873	CTTGTCTGAACCCCGAGAGACCCCATGTTTCAGGTCATCGGAGACCTCTTCTTAAGAGGT	932
QY	998	GATCAAAAGTTTGGCACAGACCAATCTATGAGGGCGCACATTTCAATGATGACAGCC	1057
Db	933	GATCAAGAGTTTGGCACAGACCAATCTATGAGGGCGCACATTTCAAGATGATGACCC	992
QY	1058	AACCTTCTTCAGAGCCCTCTACTTGGCGGACCCACACTGCCGCTCTATGAGCCATGAC	1117
Db	993	GCTGTCTTCAGACCCCGGCTACTCTTCAGAGGTCAGACGTGCGTTCAGGTTCATGAC	1052
QY	1118	TGCATGATATCTAGAGCTGTGTGGTGTCCAGGCTGGCTCTTTCACAGACCAACCGCA	1177
Db	1053	CGAGACGAGCCCAAGAGCAGTGTGCTATGAGGAGCTGCTCTTTCACACCAACCGCA	1112
QY	1178	GTTCTGGGGGCCGCCCAAGATCAAGGCTGTGTGGAGAGCTGTGCCCGGCGGCTCCT	1237
Db	1113	CTTCTGGGAGCGCGAGAGTGCAGGCGCTGTCTGCACCGCGTGCCTTCGCGAGATGAT	1172
QY	1238	GGTGTGACCTGTTGTGTGAGAGCAGCGCTGTATACCCGCACTGCTCCTTCCAGG	1297
Db	1173	CGTTCGACCTCTTGGGAGATCCAGGCGCGTGTACCAATGACCGAGTCTTCTTAGCG	1232
QY	1298	CCAGCCCTCACTTGTGATGCTGCACAACTTTGGGGGAAACCAATGCTCTTTTGAGC	1357
Db	1233	GCAGCCCTCACTGTGTGATGCTGCACAACTTTGGGGGAAACCAACGGGCTTTTGAGC	1292
QY	1358	CCTAGAGCTGTGAACCGAGGCTCCAGAAAGCTGCCGCTCTTTCCTCAACTCCACATGCT	1417
Db	1293	CGTGAAGCTGTATCAACCAACGGGCTTTTGCAACGGGCGGCTTCCCAACTCCACATGCT	1352
QY	1418	AGGCACGGGACATGGGCCCGCGAGGAGATCAGCCAGAACGAAGTGTCTATTCCTCATGAC	1477
Db	1353	GGGACACGGGCTGTGCCCGAGGGGATGAGAGAAACATGTGTATACAGACTGATGA	1412
QY	1478	TGAGCTGGGCTGGCGAAGAGACCAATGACAGATTTGGACGCTGGGTGACCAAGCTTGGC	1537
Db	1413	CGAATGGGCTGGCGCCAGAGGCC---CCTGCACCTCCCGAGCTGGTGTGCCCGCTATGAC	1468
QY	1558	CGCCCGGAGTATGGGATCTCCCAACCGGACCGACAGGGGAGGATGAGAGCTATGCTCGG	1597
Db	1470	GGAGGCGGCTACGGAGGCCCGAGAGCTGTGACGCGGACGCGCTTGGACAGTGTCTTCGG	1522
QY	1598	GAGTGTGACACTGCTCGGGGAGGCTTGACAGGGGCCAATATCGTACCGCGTGTGAG	1657

Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	CAGCGTGTAAACAATGACAC--GGGGGTGTGGTCAACCAACCAACCCGACCCGGTGGTCCG	GGGGCGGTCCTTACAGATGAATATACAGCATGTGTGATCAACAACCGATGTGATGTGTGGAGGC	CCGGCCCTCCCGCGGACATGAGACACCGAGAGTCTGTGTATCAACAAGAGCCAGTGTATGAGAGGC	CTGGGGGTGTGTGTCTCAATCTGTCTCTCTCTCTGTGGCCACACGACCCCGCTTCCGTATGCA	CTGGGGCTGTGTGTGTGAGCGCGCGCGAGACTGGAGCTTCCAGCCCTTCCAGCTTCCGTATGCA	CTGTGTGACCTCACTCGGACGAGCGAGTGTGACAGAGTGTGTGAGTGTGTATGTATGTGAGAGGC	CCTGGCGGACGCTCAAGCGGACGCGGCTCGCTCCGAGAGTGTCTCAAGCGCGCGGTGTGTGTGT	CTATGAGCGT	CTATGAGCGT	CTATGAGCGT	CTATGAGCGT	CAGCTGT	CCGCTGT	GCAAGAACAGCCGCTTACCACTGACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	GCTGAACCCCGGAGACCAAGAGT	CAACAAGCAGT	CAACAAGCAGT	GAAGCGCTGT	CTGT	AAATGTCTTCCCACTGT	GGCGCTTCTTCCAGGTATGAGAGAGGCTTCACTTACCAACAAGAGGCTTCACTTCCAGAGCC	GGGAGGAGACACTGT	TGTGGGAGACACGCTGT
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187
Db	1530	1558	1587	1718	1647	1778	1707	1838	1767	1898	1827	1958	1887	2018	1947	2078	2007	2138	2067	2198	2127	2258	2187</

REFERENCE 3 (bases 1 to 7721)  
 AUTHORS Li, H.H., Zhao, K.W. and Neufeld, E.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-MAY-1997) Biol. Chem., UCLA, 10833 Le Conte Ave., Los Angeles, CA 90095-1337, USA  
 FEATURES  
 SOURCE Location/Qualifiers  
 1. 7721  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="129sv"  
 /db\_xref="taxon:10090"  
 /chromosome="11"  
 join(206..582,1160..1307,1966..2112,2287..2372,  
 3918..4174,6311..7515)  
 /product="alpha-N-acetylglucosaminidase"  
 join(206..582,1160..1307,1966..2112,2287..2372,3918..4174,  
 6311..7515)  
 /codon\_start=1  
 /product="alpha-N-acetylglucosaminidase"  
 /protein\_id="AAC28842.1"  
 /db\_xref="GI:3129361"

mRNA  
 CDS  
 join(206..582,1160..1307,1966..2112,2287..2372,3918..4174,  
 6311..7515)  
 /codon\_start=1  
 /product="alpha-N-acetylglucosaminidase"  
 /protein\_id="AAC28842.1"  
 /db\_xref="GI:3129361"

BASE COUNT 1735 a 2050 c 2056 g 1880 t  
 ORIGIN

Query Match 34.0%; Score 874.8; DB 10; Length 7721;  
 Beef Local Similarity 82.1%; Pred. No. 1.7e-107;  
 Matches 1021; Conservative 0; Mismatches 217; Indels 6; Gaps 1;

1116 ACTGAGTGTACTAGAGCTGTGTGCTGCTCCAGAGCTGCTTTCAGACACAGCCG 1175  
 6304 ACTGTAGTGAACCTGATGCTGTGCTGCTCCAGAGCTGCTTTCAGACACAGCC 6363  
 1176 CAGTCTGGGGCCCCCAGATCAGGGCTGTGCTGGAGAGCTGCCCCGGCGGCTC 1235  
 6364 CAGTCTGGGGCCCCCAGATCAGGGCTGTGCTGGAGAGCTGCCCCGGCGGCTC 6423  
 1236 CTGCTTCTGAGACTTGTGCTAGAGCAGCTGTGTATACCCGACAGCTCCTTCAG 1285  
 6424 CTGCTTCTGAGACTTGTGCTAGAGCAGCTGTGTATACCCGACAGCTCCTTCAG 6483  
 1296 GGCAGCCCTTCATCTGTGATGCTGTGCAACAATTGGGGGAAACATGATCTTTTGA 1355  
 6484 GGCAGCCCTTCATCTGTGATGCTGTGCAACAATTGGGGGAAACATGATCTTTTGA 6543  
 1356 GGCCTAGAGGCTGTGAACGAGAGCCAGAGCTGCCCTCTTCCCAACATCCACATG 1415  
 6544 GGCCTAGAGGCTGTGAACGAGAGCCCAAGGAGCTGCCCTCTTCCCAACATCCACATG 6603  
 1416 GTAGGACGGGATGAGCCGCCAGAGGATCAGCCAGAAAGAAAGTGTATTCCTCATG 1475  
 6604 GTAGGACGGGATGAGCCGCCAGAGGATCAGCCAGAAAGAAAGTGTATTCCTCATG 6663  
 1476 GGTGAGCTGGGCTGGGAGAAAGACCAAGGACGATTTGGAGAGCTGGGTACACAGCTT 1535  
 6664 GGTGAGCTGGGCTGGGAGAAAGACCTGTACCAAGTGTGAAGGCTGGGTACACAGCTT 6723  
 1536 GCCGCGCGGCGGTATGGGCTTCCACCGGACGAGAGGGGACGCTGGAGGCTATGCTC 1595  
 6724 GCCATCCGCGGATAGGGGCTTCCACCGCTGATGCGGTGGAGGCTTCTTACTC 6783

1596 CCGAGTGTGTACACTGCTCCGGGAGAGCTGACAGGGCCACATGTAAGCCGCTGCTC 1655  
 6784 AGAAGTGTATCAACTGCTCTGGGAGGCGGTGATGAGGAGCAAAATGAAAGCCGTTGGTC 6843  
 1656 AAGCGGCGGCTCCCTCAAGATGATACCAAGATCTGTTCAACCCATCGATGTTGAG 1715  
 6844 AAGCGGCGGCTCCCTCAAGATGATACCGGCTGCTGTTCAACAGATCAAGATGTTGAG 6903  
 1716 GCTGAGCGGCTGCTGCTCAATCTGCTCCCTCCAGCCAGCCGCTTCGCTAC 1775  
 6904 GCTTGGCGACTGCTGTTAAGACTGCTGCCAAACCTGACCAAGCCAGCCGCTTCGCTAT 6963  
 1776 GACTGCTGACTCTCACTGTGGAGAGGATGAGAGGCTGTGTGATGATGATGATGAG 1835  
 6964 GACTGCTGACTCTCACTGTGGAGAGGATGAGAGGCTGTGTGATGATGATGATGAG 7023  
 1836 GCAGAGAGCGCTTACCTGAGCAAGAGCTGAGCTCCCTGTTGAGGAGGCTGCTG 1895  
 7024 GCAGAGAGCGCTTACCTGAGCAAGAGCTGATGATGATGATGATGATGATGATGATG 7083  
 1896 GCTTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1955-  
 7084 GCTTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7143  
 1956 GCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2015  
 7144 GCTTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7203  
 2016 GAGCAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2075  
 7204 GAGCAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7263  
 2076 GCGCAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2135  
 7264 GCGCAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7323  
 2136 CTGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2195  
 7324 TTGGGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7383  
 2196 AAAATGCTTCTCACTGAGAGAGGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 2255  
 7384 AAAATGCTTCTCACTGAGAGAGGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 7443  
 2256 CCGCAGAGAGAGAGAGAGGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2315  
 7444 CCGCAGAGAGAGAGAGAGGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7497  
 2316 GTGGCGGCTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2355  
 7498 CAGCTGAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7541

RESULT 15  
 AF63242 10938 bp DNA linear ROD 02-MAY-2002  
 LOCUS AF63242  
 DEFINITION Mus musculus N-acetylglucosaminidase (Naglu) and  
 17-beta-hydroxysteroid dehydrogenase (Hsd17b1) genes, complete cds.  
 ACCESSION AF63242  
 VERSION AF63242.1 GI:20385159  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 10938)  
 2 (bases 1 to 10938)  
 3 (bases 1 to 10938)  
 4 (bases 1 to 10938)  
 5 (bases 1 to 10938)  
 6 (bases 1 to 10938)  
 7 (bases 1 to 10938)  
 8 (bases 1 to 10938)  
 9 (bases 1 to 10938)  
 10 (bases 1 to 10938)  
 11 (bases 1 to 10938)  
 12 (bases 1 to 10938)  
 13 (bases 1 to 10938)  
 14 (bases 1 to 10938)  
 15 (bases 1 to 10938)  
 16 (bases 1 to 10938)  
 17 (bases 1 to 10938)  
 18 (bases 1 to 10938)  
 19 (bases 1 to 10938)  
 20 (bases 1 to 10938)  
 21 (bases 1 to 10938)  
 22 (bases 1 to 10938)  
 23 (bases 1 to 10938)  
 24 (bases 1 to 10938)  
 25 (bases 1 to 10938)  
 26 (bases 1 to 10938)  
 27 (bases 1 to 10938)  
 28 (bases 1 to 10938)  
 29 (bases 1 to 10938)  
 30 (bases 1 to 10938)  
 31 (bases 1 to 10938)  
 32 (bases 1 to 10938)  
 33 (bases 1 to 10938)  
 34 (bases 1 to 10938)  
 35 (bases 1 to 10938)  
 36 (bases 1 to 10938)  
 37 (bases 1 to 10938)  
 38 (bases 1 to 10938)  
 39 (bases 1 to 10938)  
 40 (bases 1 to 10938)  
 41 (bases 1 to 10938)  
 42 (bases 1 to 10938)  
 43 (bases 1 to 10938)  
 44 (bases 1 to 10938)  
 45 (bases 1 to 10938)  
 46 (bases 1 to 10938)  
 47 (bases 1 to 10938)  
 48 (bases 1 to 10938)  
 49 (bases 1 to 10938)  
 50 (bases 1 to 10938)  
 51 (bases 1 to 10938)  
 52 (bases 1 to 10938)  
 53 (bases 1 to 10938)  
 54 (bases 1 to 10938)  
 55 (bases 1 to 10938)  
 56 (bases 1 to 10938)  
 57 (bases 1 to 10938)  
 58 (bases 1 to 10938)  
 59 (bases 1 to 10938)  
 60 (bases 1 to 10938)  
 61 (bases 1 to 10938)  
 62 (bases 1 to 10938)  
 63 (bases 1 to 10938)  
 64 (bases 1 to 10938)  
 65 (bases 1 to 10938)  
 66 (bases 1 to 10938)  
 67 (bases 1 to 10938)  
 68 (bases 1 to 10938)  
 69 (bases 1 to 10938)  
 70 (bases 1 to 10938)  
 71 (bases 1 to 10938)  
 72 (bases 1 to 10938)  
 73 (bases 1 to 10938)  
 74 (bases 1 to 10938)  
 75 (bases 1 to 10938)  
 76 (bases 1 to 10938)  
 77 (bases 1 to 10938)  
 78 (bases 1 to 10938)  
 79 (bases 1 to 10938)  
 80 (bases 1 to 10938)  
 81 (bases 1 to 10938)  
 82 (bases 1 to 10938)  
 83 (bases 1 to 10938)  
 84 (bases 1 to 10938)  
 85 (bases 1 to 10938)  
 86 (bases 1 to 10938)  
 87 (bases 1 to 10938)  
 88 (bases 1 to 10938)  
 89 (bases 1 to 10938)  
 90 (bases 1 to 10938)  
 91 (bases 1 to 10938)  
 92 (bases 1 to 10938)  
 93 (bases 1 to 10938)  
 94 (bases 1 to 10938)  
 95 (bases 1 to 10938)  
 96 (bases 1 to 10938)  
 97 (bases 1 to 10938)  
 98 (bases 1 to 10938)  
 99 (bases 1 to 10938)  
 100 (bases 1 to 10938)  
 101 (bases 1 to 10938)  
 102 (bases 1 to 10938)  
 103 (bases 1 to 10938)  
 104 (bases 1 to 10938)  
 105 (bases 1 to 10938)  
 106 (bases 1 to 10938)  
 107 (bases 1 to 10938)  
 108 (bases 1 to 10938)  
 109 (bases 1 to 10938)  
 110 (bases 1 to 10938)  
 111 (bases 1 to 10938)  
 112 (bases 1 to 10938)  
 113 (bases 1 to 10938)  
 114 (bases 1 to 10938)  
 115 (bases 1 to 10938)  
 116 (bases 1 to 10938)  
 117 (bases 1 to 10938)  
 118 (bases 1 to 10938)  
 119 (bases 1 to 10938)  
 120 (bases 1 to 10938)  
 121 (bases 1 to 10938)  
 122 (bases 1 to 10938)  
 123 (bases 1 to 10938)  
 124 (bases 1 to 10938)  
 125 (bases 1 to 10938)  
 126 (bases 1 to 10938)  
 127 (bases 1 to 10938)  
 128 (bases 1 to 10938)  
 129 (bases 1 to 10938)  
 130 (bases 1 to 10938)  
 131 (bases 1 to 10938)  
 132 (bases 1 to 10938)  
 133 (bases 1 to 10938)  
 134 (bases 1 to 10938)  
 135 (bases 1 to 10938)  
 136 (bases 1 to 10938)  
 137 (bases 1 to 10938)  
 138 (bases 1 to 10938)  
 139 (bases 1 to 10938)  
 140 (bases 1 to 10938)  
 141 (bases 1 to 10938)  
 142 (bases 1 to 10938)  
 143 (bases 1 to 10938)  
 144 (bases 1 to 10938)  
 145 (bases 1 to 10938)  
 146 (bases 1 to 10938)  
 147 (bases 1 to 10938)  
 148 (bases 1 to 10938)  
 149 (bases 1 to 10938)  
 150 (bases 1 to 10938)  
 151 (bases 1 to 10938)  
 152 (bases 1 to 10938)  
 153 (bases 1 to 10938)  
 154 (bases 1 to 10938)  
 155 (bases 1 to 10938)  
 156 (bases 1 to 10938)  
 157 (bases 1 to 10938)  
 158 (bases 1 to 10938)  
 159 (bases 1 to 10938)  
 160 (bases 1 to 10938)  
 161 (bases 1 to 10938)  
 162 (bases 1 to 10938)  
 163 (bases 1 to 10938)  
 164 (bases 1 to 10938)  
 165 (bases 1 to 10938)  
 166 (bases 1 to 10938)  
 167 (bases 1 to 10938)  
 168 (bases 1 to 10938)  
 169 (bases 1 to 10938)  
 170 (bases 1 to 10938)  
 171 (bases 1 to 10938)  
 172 (bases 1 to 10938)  
 173 (bases 1 to 10938)  
 174 (bases 1 to 10938)  
 175 (bases 1 to 10938)  
 176 (bases 1 to 10938)  
 177 (bases 1 to 10938)  
 178 (bases 1 to 10938)  
 179 (bases 1 to 10938)  
 180 (bases 1 to 10938)  
 181 (bases 1 to 10938)  
 182 (bases 1 to 10938)  
 183 (bases 1 to 10938)  
 184 (bases 1 to 10938)  
 185 (bases 1 to 10938)  
 186 (bases 1 to 10938)  
 187 (bases 1 to 10938)  
 188 (bases 1 to 10938)  
 189 (bases 1 to 10938)  
 190 (bases 1 to 10938)  
 191 (bases 1 to 10938)  
 192 (bases 1 to 10938)  
 193 (bases 1 to 10938)  
 194 (bases 1 to 10938)  
 195 (bases 1 to 10938)  
 196 (bases 1 to 10938)  
 197 (bases 1 to 10938)  
 198 (bases 1 to 10938)  
 199 (bases 1 to 10938)  
 200 (bases 1 to 10938)  
 201 (bases 1 to 10938)  
 202 (bases 1 to 10938)  
 203 (bases 1 to 10938)  
 204 (bases 1 to 10938)  
 205 (bases 1 to 10938)  
 206 (bases 1 to 10938)  
 207 (bases 1 to 10938)  
 208 (bases 1 to 10938)  
 209 (bases 1 to 10938)  
 210 (bases 1 to 10938)  
 211 (bases 1 to 10938)  
 212 (bases 1 to 10938)  
 213 (bases 1 to 10938)  
 214 (bases 1 to 10938)  
 215 (bases 1 to 10938)  
 216 (bases 1 to 10938)  
 217 (bases 1 to 10938)  
 218 (bases 1 to 10938)  
 219 (bases 1 to 10938)  
 220 (bases 1 to 10938)  
 221 (bases 1 to 10938)  
 222 (bases 1 to 10938)  
 223 (bases 1 to 10938)  
 224 (bases 1 to 10938)  
 225 (bases 1 to 10938)  
 226 (bases 1 to 10938)  
 227 (bases 1 to 10938)  
 228 (bases 1 to 10938)  
 229 (bases 1 to 10938)  
 230 (bases 1 to 10938)  
 231 (bases 1 to 10938)  
 232 (bases 1 to 10938)  
 233 (bases 1 to 10938)  
 234 (bases 1 to 10938)  
 235 (bases 1 to 10938)  
 236 (bases 1 to 10938)  
 237 (bases 1 to 10938)  
 238 (bases 1 to 10938)  
 239 (bases 1 to 10938)  
 240 (bases 1 to 10938)  
 241 (bases 1 to 10938)  
 242 (bases 1 to 10938)  
 243 (bases 1 to 10938)  
 244 (bases 1 to 10938)  
 245 (bases 1 to 10938)  
 246 (bases 1 to 10938)  
 247 (bases 1 to 10938)  
 248 (bases 1 to 10938)  
 249 (bases 1 to 10938)  
 250 (bases 1 to 10938)  
 251 (bases 1 to 10938)  
 252 (bases 1 to 10938)  
 253 (bases 1 to 10938)  
 254 (bases 1 to 10938)  
 255 (bases 1 to 10938)  
 256 (bases 1 to 10938)  
 257 (bases 1 to 10938)  
 258 (bases 1 to 10938)  
 259 (bases 1 to 10938)  
 260 (bases 1 to 10938)  
 261 (bases 1 to 10938)  
 262 (bases 1 to 10938)  
 263 (bases 1 to 10938)  
 264 (bases 1 to 10938)  
 265 (bases 1 to 10938)  
 266 (bases 1 to 10938)  
 267 (bases 1 to 10938)  
 268 (bases 1 to 10938)  
 269 (bases 1 to 10938)  
 270 (bases 1 to 10938)  
 271 (bases 1 to 10938)  
 272 (bases 1 to 10938)  
 273 (bases 1 to 10938)  
 274 (bases 1 to 10938)  
 275 (bases 1 to 10938)  
 276 (bases 1 to 10938)  
 277 (bases 1 to 10938)  
 278 (bases 1 to 10938)  
 279 (bases 1 to 10938)  
 280 (bases 1 to 10938)  
 281 (bases 1 to 10938)  
 282 (bases 1 to 10938)  
 283 (bases 1 to 10938)  
 284 (bases 1 to 10938)  
 285 (bases 1 to 10938)  
 286 (bases 1 to 10938)  
 287 (bases 1 to 10938)  
 288 (bases 1 to 10938)  
 289 (bases 1 to 10938)  
 290 (bases 1 to 10938)  
 291 (bases 1 to 10938)  
 292 (bases 1 to 10938)  
 293 (bases 1 to 10938)  
 294 (bases 1 to 10938)  
 295 (bases 1 to 10938)  
 296 (bases 1 to 10938)  
 297 (bases 1 to 10938)  
 298 (bases 1 to 10938)  
 299 (bases 1 to 10938)  
 300 (bases 1 to 10938)  
 301 (bases 1 to 10938)  
 302 (bases 1 to 10938)  
 303 (bases 1 to 10938)  
 304 (bases 1 to 10938)  
 305 (bases 1 to 10938)  
 306 (bases 1 to 10938)  
 307 (bases 1 to 10938)  
 308 (bases 1 to 10938)  
 309 (bases 1 to 10938)  
 310 (bases 1 to 10938)  
 311 (bases 1 to 10938)  
 312 (bases 1 to 10938)  
 313 (bases 1 to 10938)  
 314 (bases 1 to 10938)  
 315 (bases 1 to 10938)  
 316 (bases 1 to 10938)  
 317 (bases 1 to 10938)  
 318 (bases 1 to 10938)  
 319 (bases 1 to 10938)  
 320 (bases 1 to 10938)  
 321 (bases 1 to 10938)  
 322 (bases 1 to 10938)  
 323 (bases 1 to 10938)  
 324 (bases 1 to 10938)  
 325 (bases 1 to 10938)  
 326 (bases 1 to 10938)  
 327 (bases 1 to 10938)  
 328 (bases 1 to 10938)  
 329 (bases 1 to 10938)  
 330 (bases 1 to 10938)  
 331 (bases 1 to 10938)  
 332 (bases 1 to 10938)  
 333 (bases 1 to 10938)  
 334 (bases 1 to 10938)  
 335 (bases 1 to 10938)  
 336 (bases 1 to 10938)  
 337 (bases 1 to 10938)  
 338 (bases 1 to 10938)  
 339 (bases 1 to 10938)  
 340 (bases 1 to 10938)  
 341 (bases 1 to 10938)  
 342 (bases 1 to 10938)  
 343 (bases 1 to 10938)  
 344 (bases 1 to 10938)  
 345 (bases 1 to 10938)  
 346 (bases 1 to 10938)  
 347 (bases 1 to 10938)  
 348 (bases 1 to 10938)  
 349 (bases 1 to 10938)  
 350 (bases 1 to 10938)  
 351 (bases 1 to 10938)  
 352 (bases 1 to 10938)  
 353 (bases 1 to 10938)  
 354 (bases 1 to 10938)  
 355 (bases 1 to 10938)  
 356 (bases 1 to 10938)  
 357 (bases 1 to 10938)  
 358 (bases 1 to 10938)  
 359 (bases 1 to 10938)  
 360 (bases 1 to 10938)  
 361 (bases 1 to 10938)  
 362 (bases 1 to 10938)  
 363 (bases 1 to 10938)  
 364 (bases 1 to 10938)  
 365 (bases 1 to 10938)  
 366 (bases 1 to 10938)  
 367 (bases 1 to 10938)  
 368 (bases 1 to 10938)  
 369 (bases 1 to 10938)  
 370 (bases 1 to 10938)  
 371 (bases 1 to 10938)  
 372 (bases 1 to 10938)  
 373 (bases 1 to 10938)  
 374 (bases 1 to 10938)  
 375 (bases 1 to 10938)  
 376 (bases 1 to 10938)  
 377 (bases 1 to 10938)  
 378 (bases 1 to 10938)  
 379 (bases 1 to 10938)  
 380 (bases 1 to 10938)  
 381 (bases 1 to 10938)  
 382 (bases 1 to 10938)  
 383 (bases 1 to 10938)  
 384 (bases 1 to 10938)  
 385 (bases 1 to 10938)  
 386 (bases 1 to 10938)  
 387 (bases 1 to 10938)  
 388 (bases 1 to 10938)  
 389 (bases 1 to 10938)  
 390 (bases 1 to 10938)  
 391 (bases 1 to 10938)  
 392 (bases 1 to 10938)  
 393 (bases 1 to 10938)  
 394 (bases 1 to 10938)  
 395 (bases 1 to 10938)  
 396 (bases 1 to 10938)  
 397 (bases 1 to 10938)  
 398 (bases 1 to 10938)  
 399 (bases 1 to 10938)  
 400 (bases 1 to 10938)  
 401 (bases 1 to 10938)  
 402 (bases 1 to 10938)  
 403 (bases 1 to 10938)  
 404 (bases 1 to 10938)  
 405 (bases 1 to 10938)  
 406 (bases 1 to 10938)  
 407 (bases 1 to 10938)  
 408 (bases 1 to 10938)  
 409 (bases 1 to 10938)  
 410 (bases 1 to 10938)  
 411 (bases 1 to 10938)  
 412 (bases 1 to 10938)  
 413 (bases 1 to 10938)  
 414 (bases 1 to 10938)  
 415 (bases 1 to 10938)  
 416 (bases 1 to 10938)  
 417 (bases 1 to 10938)  
 418 (bases 1 to 10938)  
 419 (bases 1 to 10938)  
 420 (bases 1 to 10938)  
 421 (bases 1 to 10938)  
 422 (bases 1 to 10938)  
 423 (bases 1 to 10938)  
 424 (bases 1 to 10938)  
 425 (bases 1 to 10938)  
 426 (bases 1 to 10938)  
 427 (bases 1 to 10938)  
 428 (bases 1 to 10938)  
 429 (bases 1 to 10938)  
 430 (bases 1 to 10938)  
 431 (bases 1 to 10938)  
 432 (bases 1 to 10938)  
 433 (bases 1 to 10938)  
 434 (bases 1 to 10938)  
 435 (bases 1 to 10938)  
 436 (bases 1 to 10938)  
 437 (bases 1 to 10938)  
 438 (bases 1 to 10938)  
 439 (bases 1 to 10938)  
 440 (bases 1 to 10938)  
 441 (bases 1 to 10938)  
 442 (bases 1 to 10938)  
 443 (bases 1 to 10938)  
 444 (bases 1 to 10938)  
 445 (bases 1 to 10938)  
 446 (bases 1 to 10938)  
 447 (bases 1 to 10938)  
 448 (bases 1 to 10938)  
 449 (bases 1 to 10938)  
 450 (bases 1 to 10938)  
 451 (bases 1 to 10938)  
 452 (bases 1 to 10938)  
 453 (bases 1 to 10938)  
 454 (bases 1 to 10938)  
 455 (bases 1 to 10938)  
 456 (bases 1 to 10938)  
 457 (bases 1 to 10938)  
 458 (bases 1 to 10938)  
 459 (bases 1 to 10938)  
 460 (bases 1 to 10938)  
 461 (bases 1 to 10938)  
 462 (bases 1 to 10938)  
 463 (bases 1 to 10938)  
 464 (bases 1 to 10938)  
 465 (bases 1 to 1

GSF-National Research Center for Environment and Health,  
Ingolstaedter Landstr. 1, Neuherberg 85764, Germany

## FEATURES

**Source**

1. 10938

/organism="Mus musculus"

```
/model_type="genomic DNA"
/strain="129/SVEYTacFBr"
```

```
/db_xref="taxon:10090"
```

```
/sex="female"
```

```
/c:\issue_type="splreen"
<206> 27515
```

```
/gene="Naglu"
```

```
join(<206. .582,1160.
```

```
3918. .41/4,6311. .>/gene="Naq1u"
```

```

/product="N-acetyl-g]

```

Sun Feb 15 14:49:59 2004

us-09-836-613-1.rge

Page 24

Db 7498 CAGCTGACTCTTTGTGACAGATTAGCCATGCGAGGACCTGCT 7541

Search completed: February 13, 2004, 12:03:30  
Job time : 9404 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2004, 07:06:01 / Search time 672 Seconds

(without alignments)  
10343.829 Million cell updates/sec

Title: US-09-836-613-1

Perfect score: 2575  
Sequence: 1 CCGCGGCTTAGCCTTCGGGT.....AGTCAGCGCGCGCAATTC 2575

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2575	100.0	2575	18	AAT67163
2	1426.6	55.4	10380	18	AAT67164
3	369.4	14.3	2661	23	ABL22605
4	369.4	14.3	4740	23	ABL22604
5	305.2	11.9	706	24	ABO32798
6	305.2	11.9	706	24	ABO32799
7	304.2	11.8	710	24	ABO14172
8	304.2	11.8	710	24	ABO14173

C	9	259.8	10.1	706	24	ABO32800
C	10	259.8	10.1	706	24	ABO32801
C	11	257.2	10.0	710	24	ABO14170
C	12	257.2	10.0	710	24	ABO14171
C	13	89.2	3.5	7222	23	ABL22584
C	14	80.4	3.1	114955	20	AAK53491
C	15	71	2.8	4020	18	AAI51361
C	16	70.4	2.7	5173	18	AAT89783
C	17	70.4	2.7	5173	24	ABN96859
C	18	70.4	2.7	5175	23	ABV22988
C	19	70.4	2.7	5175	23	ABV28823
C	20	70	2.7	10144	24	ABN95670
C	21	70	2.7	114955	20	AAK53491
C	22	68.6	2.7	2561	22	AAH26500
C	23	68.2	2.6	2271	24	ABZ35021
C	24	68.2	2.6	5181	24	ABL68935
C	25	66.2	2.6	3198	20	AAK02974
C	26	64.4	2.5	1614	22	AAH26499
C	27	64.4	2.5	3957	22	AAH09686
C	28	64.4	2.5	4225	25	ABV55290
C	29	64.4	2.5	14225	22	AAH26495
C	30	64.4	2.5	154746	24	AAD25519
C	31	64.4	2.5	154746	24	AAD25519
C	32	63.6	2.5	5811	24	ABN97455
C	33	63.2	2.5	43058	24	ABL65219
C	34	63.2	2.5	43058	24	ABL65219
C	35	63.2	2.5	43058	24	ABL65219
C	36	62.8	2.4	15672	12	AAO10613
C	37	62.4	2.4	3957	22	AAH09686
C	38	62.2	2.4	4411529	22	AAI99682
C	39	61.4	2.4	1562	22	AAH48175
C	40	61.2	2.4	24379	18	AAT93095
C	41	61.2	2.4	24379	19	AAV5925
C	42	61	2.4	15355	22	AAH57405
C	43	61	2.4	15359	22	ABH08724
C	44	61	2.4	15359	23	AAH59156
C	45	61	2.4	4403765	22	AAI99683

## ALIGNMENTS

RESULT 1	
ID	AAT67163 standard; cDNA; 2575 BP.
AC	AAT67163;
XX	
DT	20-AUG-1997 (first entry)
XX	
DE	Human alpha-N-acetylglucosaminidase cDNA.
XX	
KM	Alpha-N-acetylglucosaminidase; mucopolysaccharidosis type IIIB;
KW	gene therapy; enzyme replacement therapy; diagnosis; ss.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	CDS
FT	sig_peptide
FT	mat_peptide
FT	
XX	
PN	WO9719177-A1.
XX	
PD	29-MAY-1997.
XX	
PF	22-NOV-1996; 96WO-AU00747.
XX	
PR	23-NOV-1995; 95AU-0006748.
XX	

Oligonucleotide fo  
Oligonucleotide fo  
Oligonucleotide fo  
Oligonucleotide fo  
Drosophila melanog  
Human adenosine A1  
Orf virus genomic  
Human immunodefici  
Gene #357 used to  
Human prostate exp  
Human prostate exp  
Gene #2168 used to  
Human adenosine A1  
Rabbit low density  
Human gene express  
Kidney cancer rela  
Human IL-1ra BAC c  
Human low density  
HSV-2 immediate ea  
Mouse SGA2 polypep  
Human low density  
Human herpesvirus  
Human herpesvirus  
S. macromycetis  
Gene #3953 used to  
Lung cancer relate  
Lung cancer relate  
Rhinodrin receptor  
HSV-2 immediate ea  
Mycobacterium tube  
Rice gibberellin 2  
Streptomyces fredo  
Streptomyces roseo  
Human skeletal mus  
Human tyandoline re  
DNA encoding novel  
Mycobacterium tube



PA (WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.  
 XX Anson DS, Blanch L, Hopwood JJ, Scott H, Weber B;  
 XX MPI: 1997-298114/27.  
 DR P-PSDB; AAM18017.  
 XX Nucleic acid encoding mammalian alpha-N-acetylglucosaminidase -  
 PT used for the diagnosis and treatment of mucopolysaccharidosis type  
 PT IIIB, also used in gene therapy  
 XX  
 PS Claim 7; Page 46-50; 79pp; English.  
 CC A cDNA clone (AAT67163) codes for human alpha-N-acetylglucosaminidase  
 CC (AAM18017), an enzyme that can hydrolyse the terminal alpha-N-  
 CC acetylglucosamine residues at the non-reducing terminus of fragments  
 CC of heparan sulphate and heparin. It was isolated from a human  
 CC peripheral blood leukocyte cDNA library using probes based on  
 CC isolated peptide fragments (see also AAM18018-20) of the enzyme and  
 CC by examination of the corresponding genomic gene (see also AAT67164).  
 CC The isolated cDNA, and primers/probes based on the sequence or on  
 CC or its complementary strand, can be used to investigate, diagnose  
 CC and treat alpha-N-acetylglucosaminidase deficiency, such as  
 CC patients suffering from mucopolysaccharidosis type IIIB.  
 CC Administration is by oral, i.v., i.p., enzyme replacement therapy,  
 CC gene therapy or other routes.  
 CC  
 XX Sequence 2575 BP; 440 A; 819 C; 833 G; 483 T; 0 other;  
 SQ  
 Query Match 100.0%; Score 2575; DB 18; Length 2575;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCGGGCTTAGCTTCGAGTCCACGTGACCGGAGCCGCGACGTGATGACCGCGGCG 60  
 Db 1 CCCGGGCTTAGCTTCGAGTCCACGTGACCGGAGCCGCGACGTGATGACCGCGGCG 60  
 QY 61 CCCGACCCCTTGGCGCGTCCGCGGAGCCGCGACGATGACCGGAGCGGCGTGG 120  
 Db 61 CCCGACCCCTTGGCGCGTCCGCGGAGCCGCGACGATGACCGGAGCGGCGTGG 120  
 QY 121 CCGCGGCGGTGGGGGCTCTTCTCTGCGCGGAGCGCGGAGCGCGACGAGCGCC 180  
 Db 121 CCGCGGCGGTGGGGGCTCTTCTCTGCGCGGAGCGCGGAGCGCGACGAGCGCC 180  
 QY 181 GGGAGGCGGCGCGCGTCCGCGGCGCTCGTGGCCCGGCTGCGGGCCAGGCGCGCG 240  
 Db 181 GGGAGGCGGCGCGCGTCCGCGGCGCTCGTGGCCCGGCTGCGGGCCAGGCGCGCG 240  
 QY 241 ACTTCTCCGTGTCGCTGAGCGCGCTGAGCTGCGCAAGCCGCGCTTGAACCTACAGCC 300  
 Db 241 ACTTCTCCGTGTCGCTGAGCGCGCTGAGCTGCGCAAGCCGCGCTTGAACCTACAGCC 300  
 QY 301 TGGGCGGCGGCGCGCGCGCGCTGCGGGTGCAGGCGCGGCTTGAACCTACAGCC 360  
 Db 301 TGGGCGGCGGCGCGCGCGCGCTGCGGGTGCAGGCGCGGCTTGAACCTACAGCC 360  
 QY 361 CGGGGCTGACCGCTACCTGCGGCACTTCTGTGTGCTCCACGTGAGCTGATCGGCTCTC 420  
 Db 361 CGGGGCTGACCGCTACCTGCGGCACTTCTGTGTGCTCCACGTGAGCTGATCGGCTCTC 420  
 QY 421 AGCTGCGCTGCGCGCGCACTGCGCAAGCTGCGGGGAGCTGACCGGAGCGACCGCA 480  
 Db 421 AGCTGCGCTGCGCGCGCACTGCGCAAGCTGCGGGGAGCTGACCGGAGCGACCGCA 480  
 QY 481 ACAGGTACCGCTATTACAGAAATGTGACGCAAAACTACTCTCTGTGTGTGGAGT 540  
 Db 481 ACAGGTACCGCTATTACAGAAATGTGACGCAAAACTACTCTCTGTGTGTGGAGT 540  
 QY 541 GGGCGCGCTGGAGCGAGATAGTGAATGATGAGCGCTGAATGAGCATCACTGGACTGG 600  
 Db 541 GGGCGCGCTGGAGCGAGATAGTGAATGATGAGCGCTGAATGAGCATCACTGGACTGG 600

QY 601 CCGAGCGGCGGAGGAGCCATCTGGACGCGGGTGAACCTGGCTTGGGCTGACCCAG 660  
 Db 601 CCGAGCGGCGGAGGAGCCATCTGGACGCGGGTGAACCTGGCTTGGGCTGACCCAG 660  
 QY 661 CAGAGTCAATAGTCTTCTTACTGTCTGCTCTTCTGCGCTGGGGCGAATGGCAAC 720  
 Db 661 CAGAGTCAATAGTCTTCTTACTGTCTGCTCTTCTGCGCTGGGGCGAATGGCAAC 720  
 QY 721 TGCACACCTGGAGTGGCCCGCCCTCTGCGCACATCAAGAGCTTACTCTGACG 780  
 Db 721 TGCACACCTGGAGTGGCCCGCCCTCTGCGCACATCAAGAGCTTACTCTGACG 780  
 QY 781 ACCGGTCTTGGACCAAGATGCGCTCTTGGCATGACCCAGTGTGCTGCTGATTCGCG 840  
 Db 781 ACCGGTCTTGGACCAAGATGCGCTCTTGGCATGACCCAGTGTGCTGCTGATTCGCG 840  
 QY 841 GGCATGTTCCGAGGCTGACAGAGGTGTCCTCAGGTCAATGTCAAGAGTGGCA 900  
 Db 841 GGCATGTTCCGAGGCTGACAGAGGTGTCCTCAGGTCAATGTCAAGAGTGGCA 900  
 QY 901 GTTGGGCGCATTTAATGTTCTTACTCTGCTCTTCTGCTCTGCGTCCGAGAGCCCA 960  
 Db 901 GTTGGGCGCATTTAATGTTCTTACTCTGCTCTTCTGCTCTGCGTCCGAGAGCCCA 960  
 QY 961 TATTCCTCATATCGGAGGCTCTTCTGCGAGAGCTGATCAAGAGTTGGACAGACC 1020  
 Db 961 TATTCCTCATATCGGAGGCTCTTCTGCGAGAGCTGATCAAGAGTTGGACAGACC 1020  
 QY 1021 ACATCTATGGGCGGACCATCTTCAATGAGATGACAGCCACTTCTCAGAGCCCTCTAC 1080  
 Db 1021 ACATCTATGGGCGGACCATCTTCAATGAGATGACAGCCACTTCTCAGAGCCCTCTAC 1080  
 QY 1081 TTGCGGACACCACTGCGCTGATAGAGCCATGACGAGTGTATCTAGAGCTGTGT 1140  
 Db 1081 TTGCGGACACCACTGCGCTGATAGAGCCATGACGAGTGTATCTAGAGCTGTGT 1140  
 QY 1141 GGTCTCTCAAGAGGCTGCTTCTCCAGACAGCCGAGTTCGAGGGGCGCGCCAGATCA 1200  
 Db 1141 GGTCTCTCAAGAGGCTGCTTCTCCAGACAGCCGAGTTCGAGGGGCGCGCCAGATCA 1200  
 QY 1201 GGGCTGTCTGGAGCTGTGCGCGCGTGGCGGCTCTGCTGAGACCTGTTTGTGAGA 1260  
 Db 1201 GGGCTGTCTGGAGCTGTGCGCGCGTGGCGGCTCTGCTGAGACCTGTTTGTGAGA 1260  
 QY 1261 GCGAGCTGTGATACCGGCACTGCTCTTCCAGAGGCGCAACCTCTCATCTGAGTGA 1320  
 Db 1261 GCGAGCTGTGATACCGGCACTGCTCTTCCAGAGGCGCAACCTCTCATCTGAGTGA 1320  
 QY 1321 TGCACAACTTTGGGGGAAACATGATCTTTTGGAGCCCTGAGAGCTGTGAACGAGCC 1380  
 Db 1321 TGCACAACTTTGGGGGAAACATGATCTTTTGGAGCCCTGAGAGCTGTGAACGAGCC 1380  
 QY 1381 CAGAGCTGCGCGCTCTTCCCACTGACCAATGTATGAGCAAGGCAATGCCCCAGAG 1440  
 Db 1381 CAGAGCTGCGCGCTCTTCCCACTGACCAATGTATGAGCAAGGCAATGCCCCAGAG 1440  
 QY 1441 GCATCAGCAGAACGAATGTCTATTCCTCATGAGCTGAGCTGAGCGAGAAAGAGCC 1500  
 Db 1441 GCATCAGCAGAACGAATGTCTATTCCTCATGAGCTGAGCTGAGCGAGAAAGAGCC 1500  
 QY 1501 CAGTCCAGATTTGGCAGCTGAGTGAACGATTTGCGCGCGGATGAGGAGCTCTCC 1560  
 Db 1501 CAGTCCAGATTTGGCAGCTGAGTGAACGATTTGCGCGCGGATGAGGAGCTCTCC 1560  
 QY 1561 ACCGGAAGCGAGGCGAGCGTGAAGGCTACTGCTCCGAGGTGTGAACAATGCTCCGGGG 1620  
 Db 1561 ACCGGAAGCGAGGCGAGCGTGAAGGCTACTGCTCCGAGGTGTGAACAATGCTCCGGGG 1620  
 QY 1621 AGGCTGAGAGGCGCAATGTGAGCCGCTGTGAGGCGGCTCTTACAGATGAATA 1680  
 Db 1621 AGGCTGAGAGGCGCAATGTGAGCCGCTGTGAGGCGGCTCTTACAGATGAATA 1680  
 QY 1681 CCAGATGTGATCAACCGCATCTGATGTGTTGAGGCGCTGCGGCTGTGCTCAATCTG 1740





CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2661 BP; 657 A; 653 C; 698 G; 653 T; 0 other;

Query Match 14.3%; Score 369.4; DB 23; Length 2661;  
Best Local Similarity 54.0%; Pred. No. 8,2e-57;  
Matches 901; Conservative 0; Mismatches 741; Indels 27; Gaps 6;

```

QY 491 CTATTACCAAGATATGTGACACGCAAGCTACTCTCTGTGTGTGGAGTGGCCGCTG 550
DB 774 CTATACACAGAACCTGTGCACTTGAGCTTAGCTTGTGATGATGGGGCATTTAGACAGT 833
QY 551 GGAACGAGAGATAGATGATGAGCGCTGAATGAGATGAACCTGGCACTGGCCCTGAGCGG 610
DB 834 GCGGCTGCACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 890
QY 611 CCAAGAGAGGCACTGGAGAGGAGTGTACTGTGCTTGGGCTGAGACCCAGAGAGATCA 670
DB 891 CCAAGAGAGGCACTGGAGAGGAGTGTACTGTGCTTGGGCTGAGACCCAGAGAGATCA 950
QY 671 TGAATTTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 730
DB 951 CCAAGAGAGGCACTGGAGAGGAGTGTACTGTGCTTGGGCTGAGACCCAGAGAGATCA 1010
QY 731 GGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 790
DB 1011 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1070
QY 791 GGAACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 850
DB 1071 CACTGCGCAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1130
QY 851 CGAAGCTGTGACGAGGAGTGTGCTGAGATGATGATGATGATGATGATGATGATGATG 910
DB 1131 ACCTGCTGTGAGAGGAGTGTGCTGAGATGATGATGATGATGATGATGATGATGATG 1190
QY 911 CTTTAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 970
DB 1191 GTTCCCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1250
QY 971 CATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1030
DB 1251 GATAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1310
QY 1031 GAGCGGACACTTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1090
DB 1311 CTGCGATGCTTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1370
QY 1091 CACCACTGCGCTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1150
DB 1371 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1430
QY 1151 AGGCTGCTCTTTCAGACACAGCGAGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1210
DB 1431 GAGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1487
QY 1211 GGAAGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1270
DB 1488 CACGCGCGAGACAGCTGTGCTGATTTGCTGCTGAGAGGAGGAGGAGGAGGAGG 1547
QY 1271 GTATACCGGACAGCTGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1330
DB 1548 ATACGAGTAAAGCGCTCTTCTTGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1607
QY 1331 TGGGAGAAACATGATCTTTTGAAGCCTTAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1390
DB 1608 CGGCGGACACTTGAATGATTTGATCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667
QY 1391 CGGCTCTTCCCAATCCACATGATGATGATGATGATGATGATGATGATGATGATGAT 1450
DB 1668 GCGTGTGCTTCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1727
QY 1451 GAACGAAAGTGTCTATTCTCTCATGCTGAGCTGGGAGGAGGAGGAGGAGGAGGAG 1510

```

```

DB 1728 GAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1784
QY 1511 TTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1570
DB 1785 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1844
QY 1571 AGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1630
DB 1845 GAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1904
QY 1631 GGGGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1690
DB 1905 GCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1961
QY 1691 GTACACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1750
DB 1962 GTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2021
QY 1751 GGC-----CACAGCCCGCCTTCCGCTAGACCTGCTGACCTGACCTGAGCA 1798
DB 2022 ACCGCTAGAGATTAACATGATGATGATGATGATGATGATGATGATGATGATGAT 2081
QY 1799 GGCAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1858
DB 2082 ATTCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2138
QY 1859 GAGGTGAGCTTCCGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1918
DB 2139 GCGACAGGATTTGCGGATTTGCAATTCCTTAGGCTGACATGTTGAACTGTTGATGAT 2198
QY 1919 GAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1978
DB 2199 GAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2258
QY 1979 AGCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2038
DB 2259 GCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2318
QY 2039 GACCTTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2098
DB 2319 TACGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2378
QY 2099 GGTGAGCAACTATGACACCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2147
DB 2379 GGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2427

```

#### RESULT 4

ABL22604/C  
ID ABL22604 standard; DNA; 4740 BP.

XX ABL22604;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide seq ID NO 19285.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001MO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;  
 XX MPI; 2001-65860/75.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Claim 1; SEQ ID NO 19285; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins  
 CC (AB157737-AB172072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 CC  
 XX  
 XX Sequence 4740 BP; 1418 A; 1016 C; 977 G; 1329 T; 0 other;  
 SQ  
 Query Match 14.3%; Score 369.4; DB 23; Length 4740;  
 Best Local Similarity 54.0%; Pred. No. 8.5e-57;  
 Matches 901; Conservative 0; Mismatches 741; Indels 27; Gaps 6;

491 CTATTACGAAATGTGTGACGCAAGCTACTCTTCGTGTGTGGAATGAGCCGCTG 550  
 2888 CTATCCACGAAACGTGTGCACTTGAGCTATAGCTTGTGATGTGGGCAATGAGCAATG 2829  
 551 GAGAGGAGAGATAGAGTGTGATGGCGCTGATGAGCAATCAACCTGAGCTGCTGAGAGG 610  
 2828 GGGCGCTCACTGTGAGTGTGATGCGCTGTGAGGCACTGACTGATTCG--GCCGT 2772  
 611 CCAGAGGAGCATCTGTGACGAGGAGTGTACTGCTGTGAGGCTGACCCAGGAGAGATCAA 670  
 2771 CCAGAGGAGCATTTGGGTGAAAGTGTACACGGAATGAGGCTCAGAGATGAGAAATGCA 2712  
 671 TGAAGTCTTACTGTGCTGCTCTCTGAGCTGTGGAGGGAATGAGCACTGACACTG 730  
 2711 CGAGCATCTGTGTGACAGCTTTTCAAGCTGTGAGCGCAATGAGGCAATACGAGGATG 2652  
 731 GAGTGGCCCTGTGCCCCCTCTGTGACATCAAGAGCTTTTACCTGACAGCAACCGGCTCT 790  
 2651 GGGGGAGACATGACACCGGGCATGGGAGCGTATGATTAATCTCTCCAGAGAGATTAAT 2552  
 791 GGAACAGATGGGCTCTTGGGATGACCCAGTGTGCTGCTGCAATTCGGGGCATGTTCC 850  
 2591 CACTGCCCAACCAATCTGGGATGAGCGTGGCTTGGCCCTTGTGCTCAATGTGCC 2532  
 851 CGAGGCTGTACACAGGAGTGTCCCTCAGGTCAATGTCAAGAGATGGGAGTGGGGCA 910  
 2531 AGGTGCTTGAAGCTGTGATTCGGAATCGACCTTCAATGAGAGTGCAGCATGAAACA 2472  
 911 GTTAACTGTCTCACTCTGTCTCTTCTGTGCTCGAAGAGCCCAATATTCCTCAT 970  
 2471 GTTCCCGCATGATCTGTGCTGTGTTTATTTGTGAAACCAACGAGAAATCTTTCAAGA 2412  
 971 CATGGAGAGCTCTTCTGTGCGAGAGCTATCAAAAGTTGGCAAGAGCACTATATG 1030  
 2411 GATAGCTGTGCTTTTGTGATATATCAATTCGAAATATGTTGATTCACATCTTCTT 2352  
 1031 GGCAGCACTTTCAATGAGATGACGCACTTCTCAGAGCCCTCTCACTTGGCGAGC 1090  
 2351 CTGGATCTCTTCAAGAGATGAGGACCGGTACCAAGCCGGAATATGCGCTTCAAC 2292  
 1091 CACCACTGCGCTCTATGAGGCAATGACATGATGATGAGGCTGTGTGGCTCTCA 1150  
 2291 GGGGGGCGGATTTATGATCTATGCGGCGCATGATCTCTCAGGCACTTGTGCTTCA 2232

QY 1151 AGGCTGCTCTTCAAGACACGCGGAGATTCTGAGGAGCCCGCCAGATACAGGCTGTGCT 1210  
 DB 2231 GGGCTGATGTGTGTGAAGATCC---CTTTGACACGAGCATGATGAGCTTCTCT 2175  
 QY 1211 GGGAGCTGTGCCCCGGGAGCGCTCCGTGGTTGTGACCTGTTTGTGAGAGCCAGCTGT 1270  
 DB 2174 CAGGCGGACACAGTGTGTGCAATTTTGTGTCTGTCAATCTGACAGGAGCAATTTCCGA 2115  
 QY 1271 GTATCCCGAGCTGCTCTTCCAGGAGCAAGCCCTTCAATGTGTGATGCTGTGACAACTT 1330  
 DB 2114 ATAGAGACTAACCGCTCTCACTTTGGGAGAGCCCTTATATGATGCTGATCTTCAACATTT 2055  
 QY 1331 TGGGGGAAACATGATCTTTTGGAGCCCTAGAGGCTGTGAACGAGGCCGCAAACTCTC 1390  
 DB 2054 CGGGGCACTGTGAATGTTTGAATAGGAAAGCTGTATTAATTCGGAAATGAAAGGC 1995  
 QY 1391 CCGCTCTTCCCACTCCACATGTGTAGACAGGAGCATGAGCCCGGAGGATCAGCA 1450  
 DB 1994 GCGTCGCTTCCCAAGAGAGTTGTGTGGCAAGGAAATCACCCGAGGGCATTTGGCCA 1935  
 QY 1451 GAACGAAGTGTATTTCTCTCATGTGCTGAGCTGTGCGGAAAGACCAATGCTCA 1510  
 DB 1934 GAATATGTATGTATTCGTTCACTTGAAGCGCGGCTGAGATA--ATACATCCCTA 1878  
 QY 1511 TTGAGAGCTGTGATGACAGCTTTCGCGCGGAGTATGAGGCTCTCCACCCGAGCC 1570  
 DB 1877 TCTGATAGCTGTATTCGAATTTCTCACAGCCGCTATGAGGTTAAGATAGGCTTT 1818  
 QY 1571 AGGGGAGCTGTGAGGCTACTGCTCGGAGTGTACAACTGCTCCGGGAGGCTTGAC 1630  
 DB 1817 GAGAGAGGCTGTGCTGCTCTTAAGAAATAGCGCTATTCCTTCGTGCTGAGAAAT 1758  
 QY 1631 GGGGCAATCTGAGGCGCTGTGTCAGGGGCGGCTGCTCAAGATTAATACAGATTCG 1690  
 DB 1757 GCGGGACATATATATGATGACGCCA--GCTTCTTCAACCAAGAACCTTCACTTG 1701  
 QY 1691 GTACAAACCATCTGATGTGTTTGAAGCTGTGCGGCTGTGCTCACTGTCTCCCTCT 1750  
 DB 1700 GTACAAATGCAATGTGAGTGTGATGCTGTGCACTTGTGTGACCTTCAGAGCATCAT 1641  
 QY 1751 GGC-----CACAGCCCCGCTTCGCTACAGCTGTGAGACCTGACCTGAGCA 1798  
 DB 1640 ACCGCTAGAGATTAACAGATACAGATATACAGAACGATTTGTGTGACATACAGGCA 1581  
 QY 1799 GAGAGTGCAGAGAGCTGTGAGCTTGTACTATGAGAGGAGCAAGAGCCCTACCTAGCAA 1858  
 DB 1580 ATTCTGAGATCAGGAGCAGATCAATGTATATCAATCTAAGGTGGCCTA--TAGAAA 1524  
 QY 1859 GAGAGTGGCTCCCTGTTGAGGAGCTGTGAGGCTGTGAGCTATGAGTGTGCGGCACT 1918  
 DB 1523 GCAACAGGTTTCCGTTTGAAATTCCTTAGCGCAAGCTGTGAAACCTGTGATGACAT 1464  
 QY 1919 GAGAGAGTGTGAGTGTGAGCAGCGCTTGTGTGAGGAGCTGTGAGAGGAGCCCG 1978  
 DB 1463 GAGAGTGTGAGTGTGAGCAGAGAGAAATTTTGTGCTGGCACTGTGCTCAGAGGCCAA 1404  
 QY 1979 AGAGGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2038  
 DB 1403 GCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1344  
 QY 2039 GACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2098  
 DB 1343 TACCGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1284  
 QY 2099 GGTGAGCACTACTACACCCCTGTGCTGCGGCTTTTCTGTGAGGAGGCTG 2147  
 DB 1283 GTTGAAGCACTACTATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1235

RESULT 5  
 AB032798  
 ID AB032798 standard; DNA; 706 BP.  
 XX

AC AB032798;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19389.  
 XX  
 KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KM drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KM SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 EN WO200218632-A2.  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 PA (EPIC-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 DR WPI; 2002-371829/40.  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 CC  
 CC Sequence 706 BP; 76 A; 93 C; 276 G; 261 T; 0 other;  
 XX  
 SQ  
 Query Match 11.8%; Score 305.2; DB 24; Length 706;  
 Best Local Similarity 77.8%; Pred. No. 2e-45;  
 Matches 381; Conservative 0; Mismatches 108; Indels 1; Gaps 1;  
 QY 3 CGGGCTTAGCCTTCGGGTCCACGTGGCGGAGCGCCGACGTGATGAGCGCGGCGCC 62  
 DB 95 CGGGTTAGTTTTCGGGTTACGTGTCGGA-GTCGGTATGATTTGAGCGCGGTCGTT 153  
 QY 63 CCACCCCTGGCGGTCCGGGACCGGACGACGACGACGACGACGACGACGACGACGAC 122  
 DB 154 TTATTTTTCGTCCTCCCGGATTCGATGAGTTGAGTTGAGGCGGCGGCGGTC 213  
 QY 123 GCGCGGTGGGCGGCTCTTCCTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 182  
 DB 214 GCGCGGTGGGCGGCTCTTCCTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 273

QY 183 GAGCGCGCGCGCGGTGCGGCGGCTGTCGCGCCGCTGCTGCGGCGAGGCCCGCGCGAC 242  
 DB 274 GAGCGCGCGCGGTGCGGCGGCTGTCGCGCGGCTGTCGCGGCTGTCGCGGCTGTCG 333  
 QY 243 TTCTCCGTGTCGTGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302  
 DB 334 TTTTTCGTCGTGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 393  
 QY 303 GCG 362  
 DB 394 GCG 453  
 QY 363 GCGCTGACCGCTACCTGCGCGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 422  
 DB 454 GCGTTGATCGTTATTTGCGGATTTTGTGTTGATGCTGCTGCTGCTGCTGCTGCT 513  
 QY 423 TTGCGCTGCG 482  
 DB 514 TTGCGTTGTCGCGGTTATTTGATGCTGCTGCGGCGGATGATGAGGTTACGTTTAT 573  
 QY 483 AGGTACCGCT 492  
 DB 574 AGGTACGTT 583

RESULT 6  
 AB032799/c  
 ID AB032799 standard; DNA; 706 BP.  
 XX  
 XX AB032799;  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19390.  
 XX  
 KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KM drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KM SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 EN WO200218632-A2.  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 PA (EPIC-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 DR WPI; 2002-371829/40.  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method

is used (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's) and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. CC AAG1310-AB054121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in CC the disclosure of the invention.

SQ Sequence 706 BP; 261 A; 276 C; 93 G; 76 T; 0 other;

Query Match	11.9%;	Score 305.2;	DB 24;	Length 706;
Best Local Similarity	77.8%;	Pred. No. 2e-45;		
Matches 381; Conservative	0;	Mismatches 108;	Indels 1;	Gaps 1.

QY	CGGGCTTAAGCTT	CGGGTCCAGGTGACCGGAGCCGGAGCTGATGGACCGGGCCGCC	62
Db	612	CGGGTTAAGTTTTCGGGTTTACGTGTCCGA-GTGGGTAAGTGAATTGACCGGGTGGT	554
QY	63	CCACCCCTTGGCCGTGCGGGACCCGCAAGACTGAAACCATGAAAGCCGTGTGCGTGGCC	1222
Db	553	TTATTTTTCGTGTCGCGGGAATTCGTAGATTGAAATTAGAGCCGTTGCGCTGTGTC	454
QY	123	CGCGCGGTGGGGGATCCCTCTCCGAGCGGGGGCGGGGGGCGGGGACGAGCGAGCGCCGG	182
Db	493	GCGCGGTGGGGGTTTTTTTTTTTGTGCGGGGTGGGGGGCGCGTGAAGGAGTTCGG	424
QY	183	GAGCGCGCGCCGTGCGGGCGCTCGTGCCCGGCTGTGGGCGCAAGCCCGCGAGC	242
Db	433	GAGCGCGCGGTGCGGCGGGCGTTCGTTGCTGTGTGGGGTTAGGTTTGCGCGTGCAT	374
QY	243	TTTCCCGTGTGCGGTGGAAGCGCGCTCGGTGCGCAAGCGGGTGTGGAACCTACAGCTG	302
Db	373	TTTTTCGTGTGCGGTGGAAGCGCGTTCGTGTGTGAAGTCGGTTTGGAATTTAAAGTTTG	314
QY	303	GCGCGCGCGCGCGCGCGCGCGCTGTGCGGGTGCACGCTCCAGCGGCGTGCGCCCGCG	362
Db	313	GCGCGCGCGCGCGCGCGCGCGCGTGCGGGTTCGCGGGTTCAGCGGCGTGCCTGTCGCG	254
QY	363	GGGGTGACACCGTACCTCGCGGCACTTCGTGTGCTGCAGAGTGGCCCTGTCGGGCTCAG	422
Db	253	GGGTGTGATGATTAATTGCGCAATTTTGTGTGTAAAGGTGATTTGGTTCGTTTTAG	194
QY	423	CTGCGCTGCGCGGCGCACTGCGAGCGCTGCGCGGGGAGCTGACCGAGCCAGCCCAAC	482
Db	193	TTTGCTTTGTGCGCGGTATTGTTAAGTCGTGCTGGGGGAGTTGATGAGAGTTACGTTTAAT	134
QY	483	AGGTACCGCT	492
Db	133	AGGTATGTT	124

RESULT 7  
ABQ14172  
ID: ABQ14172 standard; DNA; 710 BP

AC ABQ14172;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 763.

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; ar.

KW gastrointestinal; respiratory system; single nucleotide polymorphism

XX OS Homo sapiens.

XX WO200218632-A2  
PN



PD 07-MAR-2002

PF 01-SEP-2001; 2001WO-EP10074.  
XY

PR	01-SEP-2000; 2000DE-1043826.
DP	05-SEP-2000; 2000DE-1044543.

XX (EPTG-) EPTGENOMICS AG.

XX	Olek A.	Pienbrock C.
PI		

XX  
DR WPT: 2002-371829/40.

AA  
PT  
Determining the degree

PT amplicons from chemically treated DNA - selective hybridization of  
 P1 for diagnosis and prognosis, comprises

PS Claim 12; 56pp + Sequence Listing; 56pp; German

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-GGG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridized to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridization to both classes is determined from the label on the amplicon. From the ratio of labels hybridized to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. AB01310-AB054121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

50 Sequence 710 BP; 78 A; 91 C; 276 G; 262 T; 3 other,

Query Match	11.8%	Score 304.2;	DB 24;	Length 710;
Best Local Similarity	77.4%	Pred. No. 3.1e-45;		
Matches 380;	Conservative	0;	Mismatches 110;	Indels 1;
			Gaps	1

QY	3	CGGCGTTACCCCTTGGGCTCCAGTGGCCGGAGGCGCGAGCTGATTTGAGACCGCGGGCCGCC	62
Db	187	CGGATTAGTTTTCGGGTTTACGTGTGGAGGTCGGATTTGATTTGACCGGGTCTT	246
QY	63	CCACCCCTCGCCCGTGGCGGAGCCCGCAGACTGAGCATGAGAGCGGTGGCCCTGGCC	122
Db	247	TTATTTTGTGTGTGTGGCGGATTTGTAGATTTAGATTTAGAGGCGGTGGCGGTGTCTC	306
QY	123	GCAGCGGTGGGGTCTCTTCTCTGGCCGGGCGCGGGCGCGAGGCGAGAGCCCGG	182
Db	307	GCAGCGGTGGGGTCTTTTCTTTTGGTCCGGGCGCGGGCGCGGTAGCGAGAGTTCCG	366
QY	183	GAGGCGGCGGCGGTGGGGGCGCTGTGTGACCCGGCTGTGGGGCCAGGCGCCCGCGGCGAC	242
Db	367	GAGGCGGCGGTCTGTGGGCGGCTTCTGTGTGTGGTTTGGGTTTAGTTTGGCGGTGAT	426
QY	243	TTTCCGTGTGGGTGAGGCGCGCTGTGGGTGCGCAAGCGGGCTTGGACACTTACAGCTG	302
Db	427	TTTTTGTGTGTGTGGAGGCGCTTTTGGTGTGTAGTGGGTTTGGATTTTATAGTTTG	486
QY	303	GCGGCGGCGGCGCGGCGGCGGTGGCGGTGCGGGCTCCAGCGGGTGGCGGCGCGCGG	362
Db	487	GCGGCGGCGGCGGCGGCGGCGGCGGTGGCGGGTGGCGGGCTTTTACGGGCGGTGGCGGTGTGCGG	546
QY	363	GGGCTGACCGGTACTGTGCGACATTTGTGTGGCTGCCACAGTGGCGTGTGTGCGGCTCA	422
Db	547	GGGTGTGTGTGTATTTTGGCGATTTTGTGGTGTGTACGTGGTTTGAATTCGGTTTTTA	606



QY 422 GCTGCGCTGCGCGGACAGTCCAGCCGCGGGAGAGTACCCAGAGCCACGCCCA 481  
 DB 607 GTTCCGTTTTCGCGGTTATTTAGTGTGTGTGCGGGAGTTGATCGAGTTTACGTTTAA 666  
 QY 482 CAGGTACCGCT 492  
 DB 667 TAGGTATCGTT 677

## RESULT 8

ABQ14173/c  
 ID ABQ14173 standard; DNA; 710 BP.

ABQ14173;  
 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 764.

XX Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.

XX Homo sapiens.

XX MO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CPG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridized to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridization to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridized to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 710 BP; 262 A; 276 C; 91 G; 78 T; 3 other;

XX Query Match 11.8%; Score 304.2; DB 24; Length 710;  
 XX Best Local Similarity 77.4%; Pred. No. 3.1e-45;

Matches 380; Conservative 0; Mismatches 110; Indels 1; Gaps 1;  
 QY 3 CGGGCTTACCTTCCGGTCCAGTGGCCGAGAGCCGCGAGCTGATTTGACGCGGCCGC 62  
 DB 524 CGGGTTTATGTTTTCGGGTTTACGTGTGAGTGGAGTGGATTTGACGCGGCTT 465  
 QY 63 CCACCCCTGCGCGTCCGCGGACCCGACAGACTGAGACCATGAGAGCGGTGGCGGCTC 122  
 DB 464 TTAATTTTGTGTCGTGCGGAGTTGTAGATTGATTCGAGCGGCTGGCTGATC 405  
 QY 123 GCGCGGTGCGGGGTCCTTCTCTGACCAGGCGCGGCGCGGCGGCGAGCGAGCCGCG 182  
 DB 404 GCGCGGTGCGGGGTTTCTTCTGCGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGG 345  
 QY 183 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 242  
 DB 344 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 285  
 QY 243 TTCTCGTGTGCGGAGCGGCTGCTGAGCTGCAAGCCGCGTTGAGACCTTACAGCCTG 302  
 DB 284 TTTTTCGTGTCGTGAGACGCGCTTTTGGTTGTTAGTGGGTTTGGATATTATATTGTTG 225  
 QY 303 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 362  
 DB 224 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 165  
 QY 363 GGGCTGCAACCGCTACCTGCGGACTTCTGTGACTGCCAAGCGGCTTGGACCTTACAGCCTG 421  
 DB 164 GGGTTGTATCGTTATTTGCGGATTTTGTGTTGTATGCGGTTGAGTTGAGTTTGA 105  
 QY 422 GCTGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 481  
 DB 104 GTTGGTTTGTGCGCGGTTATTTAGTGTGTGCGGGAGATTGATCGAGTTTACGTTTAA 45  
 QY 482 CAGGTACCGCT 492  
 DB 44 TAGGTATCGTT 34

## RESULT 9

ABQ32800/c  
 ID ABQ32800 standard; DNA; 706 BP.

XX ABQ32800;

XX 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19391.

XX Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.

XX Homo sapiens.

XX MO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56bp + Sequence Listing; 56bp; German.  
 XX

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's), and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ1410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 706 BP; 87 A; 93 C; 267 G; 259 T; 0 other;

XX Query Match 10.1%; Score 259.8; DB 24; Length 706;  
 XX Best Local Similarity 71.9%; Pred. No. 2.5e-37;  
 XX Matches 353; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

QY 1 CCGGGGCTTACGCTTCCGAGTGGCGGAGGCGCGGACGCTGATGAGCGGGGCG 60  
 DB 614 CCCGAACTTAACCTTCGATCCAGTAAACC-GAAACCCGAACTTAATTAACCGGAAACG 556  
 QY 61 CCCCAACCCCTGCGCGTGGCGGAGCCCGAGACTGAGACCATGAGGCGGTGGCGTGG 120  
 DB 555 CCCCAACCCCTTACCGTGGCGGAAACCGGAAACCTTAATTAACCATTAACGATTA 496  
 QY 121 CCGGGGCGGTGGGGGCTTCTTCCTGCGCGGCGGGGCGGGGCGCGGACGACGAGCCG 180  
 DB 495 CCGGACGATTAATAAATCTTCTTCCGTAACCGAAACCGAAACGCGAACAACGAAACCC 436  
 QY 181 GCGAGCGCGCGCGCTGCGGCGCTGCTGCGCGCTGCTGCGCGCGCAAGCCCGCGCGCG 240  
 DB 435 GAAAAACGACGACCGCTGAGAACGCTGTAACCCGACTTAACCAACCAACCCCGGACCG 376  
 QY 241 ACTTCTCGGTGCGGTGAGCGCGCTGCGTCCGCAACCGCGGCTTGAACACTTAACACC 300  
 DB 375 ACTTCTCGGTATCGATTAATAACCGGCTCTTAACCAACCAACCAACTTAACCTTAACAC 316  
 QY 301 TGGGCGCGCGCGCGCGCGCGCGCTGCGGCGGTGCGGCTGCAACGCGGCGTGGCGCGCG 360  
 DB 315 TAAAGACGACGACCGCGGACCGCGCTGACGAATACGAGATCCACGAAACGTAACGACCGCG 256  
 QY 361 GCGGCGTGAACCGCTTAACCTGCGGAACTTGTGCTGCGACAGTGGCGCTGCTGCTTC 420  
 DB 255 CGAAACTACACCGCTTAACCTGACGCACTTCTATTAACGATTAACCTTAACCTTAACCTTC 196  
 QY 421 AGCTGCGCGTGGCA 480  
 DB 195 AACTAGCGCTTACCGGACGACCTTAACCAACGTAACCGAAACCACTTAACCGAAACCAACGCCCA 136  
 QY 481 ACAAGTACCGC 491  
 DB 135 ACAAAATACCGC 125

RESULT 10  
 ABQ32801  
 ID ABQ32801 standard; DNA; 706 BP.  
 XX  
 AC ABQ32801;

XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19392.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
 XX SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Gnetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -

PS Claim 12; 56bp + Sequence Listing; 56bp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's), and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ1410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 706 BP; 259 A; 267 C; 93 G; 87 T; 0 other;

XX Query Match 10.1%; Score 259.8; DB 24; Length 706;  
 XX Best Local Similarity 71.9%; Pred. No. 2.5e-37;  
 XX Matches 353; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

QY 1 CCGGGGCTTACGCTTCCGAGTGGCGGAGGCGCGGACGCTGATGAGCGGGGCGCG 60  
 DB 93 CCCGAACTTAACCTTCGATCCAGTAAACC-GAAACCCGAACTTAATTAACCGGAAACG 151  
 QY 61 CCCCAACCCCTGCGCGTGGCGGAGCCCGAGACTGAGACCATGAGCGGTGGCGTGG 120  
 DB 152 CCCCAACCCCTTACCGTGGCGGAAACCGGAAACCTTAATTAACCACTTAACGATTA 211  
 QY 121 CCGGGGCGGTGGGGGCTTCTTCCTGCGCGGCGGGGCGGGGCGCGGACGCGAGAGCCG 180  
 DB 212 CCGGACGATTAATAAATCTTCTTCCGTAACCGAAACCGAAACCGGACGACGAAACCC 271  
 QY 181 GCGAGCGCGCGCGCTGCGGCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240

```

Db      272  GAAAGACGAGCGGTAAGGAGCTGTAAACCGACTACTTAAACCAACCCGCGACCG 331
Qy      241  ACTTCTCCGTGCGTGTAGAGCGGCTGTGGCTGCAAGCGGAGCTTGAGACCTAGAGCC 300
Db      332  ACTTCTCCGTATGATATTAAGGCGCTTAATCTAACCAACCGAACTTAAACCTTAAAC 391
Qy      301  TGGGCGGCGGCGGCGGCGGCGGCGGCTGTGGGCTGTGGGCTGTGGGCGGCGGCGG 360
Db      392  TAAACGACGACGACGACGCGGCTAGAAATACGCGACTCAGCAAGCTTAAACGACCGCG 451
Qy      361  CGGCGCTGACACCGCTACTGCGGCGACTTCTGTGGCTGTGGCGGCGGCTGTGGCGCTC 420
Db      452  CGAAACTACACCGCTACTGCGGCGACTTCTATTAACGACGTAACCTTAATCGACTCTC 511
Qy      421  AGCTGCGCTGTGGCGGCGCACTGCGACCGCTGTGGCGGCGGAGCTGACCGAGCCAGCCCA 480
Db      512  AACTACCGCTTACCGCGACCTACCAACCGTACCGAAATACTAACCGAAACCGACGCCCA 571
Qy      481  ACAAGTACCGC 491
Db      572  ACAATACCGC 582

```

RESULT 11  
ABQ14170/c  
ID ABQ14170 standard; DNA; 710 BP.

ABQ14170;  
12-JUL-2002 (first entry)

Oligonucleotide for detecting cytosine methylation SEQ ID NO 761.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
drug; side effect; cancer; central nervous system; cardiovascular;  
gastrointestinal; respiratory system; single nucleotide polymorphism;  
SNP; cell differentiation; de.

Homo sapiens.

WO200218632-A2.

07-MAR-2002.

01-SEP-2001; 2001WO-BP10074.

01-SEP-2000; 2000DE-1043825.

05-SEP-2000; 2000DE-1044543.

(EPig-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K, Gnetig D;

WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful  
for diagnosis and prognosis, comprises selective hybridization of  
amplicons from chemically treated DNA -

Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of  
methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
genomic sample of DNA. The sample is treated chemically to convert  
cytosine (C) but not methylated C, to uracil, then part of the genomic  
DNA that contains the target C is amplified to form a labeled amplicon.  
The amplicon is hybridized to two classes, each with at least one  
member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
and the degree of hybridization to both classes is determined from the  
label on the amplicon. From the ratio of labels hybridized to the two  
classes of oligomers, the degree of methylation is calculated. The method  
is used: (1) for diagnosis and/or prognosis of side effects of

CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
of the central nervous, cardiovascular, gastrointestinal and respiratory  
systems etc., particularly by detecting mutations or single nucleotide  
polymorphisms (SNP's); and (11) for differentiation of cell or tissue  
types and for investigating cell differentiation. The method allows the  
methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

Sequence 710 BP; 103 A; 91 C; 250 G; 263 T; 3 other;

Query Match 10.0%; Score 257.2; DB 24; Length 710;  
Best Local Similarity 71.3%; Pred. No. 7.3e-37;  
Matches 351; Conservative 0; Mismatches 140; Indels 1; Gaps 1;

```

Qy      1  CCGGCGCTTACGCTTGGGCTCCAGGTGGCGGAGCGCGGCGGAGCTTGAACGGGGCGG 60
Db      526  CCGGACTTAACCTTGCATCCAGTTAACGAAACGACAACTTAATTAACGGAAACCG 467
Qy      61  CCCCACCCCTGCGCGGTGCGGGAACCGCAGACTGAGACCATGAGAGCGGTGGGTG 120
Db      466  CCCCACCCCTTAAACCGTGCAGAAACCGCAAACTTAAACCATTAACGATTAAGATAA 407
Qy      121  CCGCGCGGTGGGGGTCTTCTTCTGCGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGG 180
Db      406  CCGGACGATTAATAATCTTCTCTTAACCGAAACGAAACGCAAAACGACGAAACCG 347
Qy      181  GAGAGCGCGCGCGCGTGTGGCGGCGCTGTGGCGCGCTGTGGCGCGCGCGCGCGCGG 240
Db      346  GAAAGACGACCGCTTACGAAACGCTCGTAACCGGACTACTTAACCAAAACCGCGACCG 287
Qy      241  ACTTCTCCGTGTGCGTGTGAGAGCGCTTGTGCTGCAAGCGGAGCTTGAACCTTACAGCC 300
Db      286  ACTTCTCCGTATGATATTAAGGCGCTTAATCTAACCAACCGAACTTAAACCTTAAAC 227
Qy      301  TGGCGGCGGCGGCGGCGCGCGCGCTGTGGGCTGTGGCGGCTGTGGCGGCGGCGCGG 360
Db      226  TAAACGACGAGGAAACGACGCGCGCTAGCAATTCGCACTCCAGCAAGTAAACGACCGCG 167
Qy      361  CGGAGCTGCAACCGTACTTGTGGCGGACTTGTGTGCTGTGCAACGCTGAGCGCTT 419
Db      166  CGAAACTACACCGCTACTACGCGACTTCTTAATCTAACGTAACCTTAATCCGACTCT 107
Qy      420  CAGCTGCGCGCTGTGGCGGCGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 479
Db      106  CACTTACGCGCTTACCGGCACTACCAACCGTACCGAAATACTTAACCGAAACCGACCGC 47
Qy      480  AACAGGTACCGC 491
Db      46  AACAAATACCGC 35

```

RESULT 12

ABQ14171  
ID ABQ14171 standard; DNA; 710 BP.

ABQ14171;

12-JUL-2002 (first entry)

Oligonucleotide for detecting cytosine methylation SEQ ID NO 762.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
drug; side effect; cancer; central nervous system; cardiovascular;  
gastrointestinal; respiratory system; single nucleotide polymorphism;  
SNP; cell differentiation; de.

Homo sapiens.

WO200218632-A2.

07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.  
 PF 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 PA (EPIC-) EPIGENOMICS AG.  
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX MPI, 2002-371829/40.  
 DR

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CGG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 710 BP; 263 A; 250 C; 91 G; 103 T; 3 other;

Query Match 10.0%; Score 257.2; DB 24; Length 710;  
 Best Local Similarity 71.3%; Pred. No. 7.3e-37;  
 Matches 351; Conservative 0; Mismatches 140; Indels 1; Gaps 1;

QY 1 CCGGGCTTACCTTCCGCTCCAGTGGCCGAGCCGAGCTGATGAGCCGGGCGG 60  
 DB 185 CCGAACTTAACCTTCGATCAAGTAACGAAACCGAACCTTAATTAACGGAAACCG 244  
 QY 61 CCGCAACCCCTGCGCGTGGCGGACCCGCAAGACTGAGACCATGAGCGGTGGCGTGG 120  
 DB 245 CCGCAACCCCTTACCGTGGCGAAACCCGCAAACTTAACCATTAACGATTAACGATTA 304  
 QY 121 CCGGGCGGTGGGGGTCTTCTTCTGCGCGGGGGCGGGGGCGGGGCGGAGGAGCC 180  
 DB 305 CCGGACACATTAATAATCTTCTCTTAACCGAAACCGAAACCGAACGAGAAACCC 364  
 QY 181 GCGAGCGCGCGCGCGTGGCGGCGCTCGTGGCGCGCGCTGCGGAGCGAGCCCGCGCG 240  
 DB 365 GAAACACGACACCGTACGAGACGCTCGTAACCGACCTTAATAACCAACCCCGGACCG 424  
 QY 241 ACTTCTCCGTGCGGTGAGCGCGCTCTGAGTGCAGACCGGGCTTGAACAACCTACAGCC 300  
 DB 425 ACTTCTCCGTATCGATTAATAACGCGCTCTTAACCTAACCAACCTTAACCAACCAAC 484  
 QY 301 TGGGGCGGGCGGGCGCGCGCGCGGCGGGTGGCGCGGCTCCAGCGGGCGTGGCGCGCG 360  
 DB 485 TAAACACGACGAGGACGCGCGCGCTGAATACGAGCTCCAGCAAGTAAACGACCGCG 544  
 QY 361 CCGGGCTGACCGCTACCTGCGGCACTTCTGTGGCTGCGCAAGTGGCGCT-GGTGGGCTCT 419  
 DB 545 CGAACTACACCGCTACCTACGCACTTCTTAACCTACGAGTAACTTAATCCGACTCT 604

QY 420 CAGTGGCGCTTGGCGGCGCACTGCGCAAGCCGTGCGGGGAGCTGACCGAGCCACCGCC 479  
 DB 605 CACTACCGCTTACCGCGAGCACTACCAACCGTACCGAAATACTAACCGAAACGACCGCC 664  
 QY 480 AACAGTACCGCG 491  
 DB 665 AACAAATACCGC 676

RESULT 13

ABL22584/c  
 ID ABL22584 standard; DNA; 7222 BP.

XX ABL22584;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 19225.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmacological; gene; ds.

OS Drosophila melanogaster.

PN MO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

DR MPI; 2001-656860/75.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PS interactions -  
 Claim 1; SEQ ID NO 19225; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA

CC sequences (AB18737-AB192072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

Sequence 7222 BP; 2085 A; 1560 C; 1634 G; 1943 T; 0 other;

Query Match 3.5%; Score 89.2; DB 23; Length 7222;  
 Best Local Similarity 54.3%; Pred. No. 7.1e-07;  
 Matches 203; Conservative 0; Mismatches 168; Indels 3; Gaps 1;

QY 1774 AGACCTGTGACCTCACTGCGAGGAGTGCAGAGCTGTGCTGACTTGAATGAGG 1833  
 DB 7145 ACGATTGGTGAATTAACAGGCAATTCGCAATGATGAGCGAGATGATGATGATCA 7086  
 QY 1834 AGGCAAGAAAGCCCTTACTGAGCAAGAGCTGGCTCTGCTGAGGCTGAGAGCGTCC 1893  
 DB 7085 ATCTAAGGTGGCTTA---TGAAGGCAAGAGTTGCGTTTCAATTCCTTAAGCTCA 7025  
 QY 1894 TGGCTATAGAGTGTGCGCGGCACTGAGCAAGGTGCTGCTAGTACAGCGGCTTCTTGC 1953

Db 7028 AGCTGTGAAAAGTGTGATGATGAGCTGATCTGGCGAGCAGAGAAATTTCTGC 6569  
 Qy 1954 TGGGACAGTGGCTAGAGAGAGCCGAGAGAGTCAAGAGCCGAGCCGATTTCT 2013  
 Db 6968 TCGGCACTGGCTGACAGAGGCCAAGCAGACCGGACACATCGGCAACAAGAACT 6909  
 Qy 2014 ACGGACAGAGAGCCGCTACCACTGACCTTTGGGGGCGCAAGGCAATCTGACT 2073  
 Db 6908 TCGGATTCATGCGAGAAACCAATTAACCGCTGGGGTCTGATGATGCAATCTGACT 6849  
 Qy 2074 ATGGCAACAAGCAGCTGGCGGGGTGGTGGCACTACTACCCCTGGCTGGCGGCTT 2133  
 Db 6848 ATGGCTGCAACATGTGCTCCGCTTGGTGAAGCACTATAGGCTTGGTGAAGCTCT 6789  
 Qy 2134 TCTGGAGGCGCTG 2147  
 Db 6788 TCCGTGAAGATGTG 6775

RESULT 14  
 AAX53491/C  
 ID AAX53491 standard; DNA; 114955 BP.  
 AC AAX53491;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE Human adenosine A1 receptor antisense oligonucleotide fragment.  
 XX  
 KW Antisense oligonucleotide; multiple target; antisense treatment;  
 KW impaired respiration; inflammation; lung disease;  
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;  
 KW acute asthma; allergy; asthma; impeded respiration;  
 KW respiratory distress syndrome; pain; cystic fibrosis;  
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;  
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;  
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;  
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
 KW prostate cancer; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN W09913886-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 17-SEP-1998; 98WO-US19419.  
 XX  
 PR 09-JUN-1998; 98US-0093972.  
 PR 17-SEP-1997; 97US-0059160.  
 XX  
 PA (UYEC-) UNIV EAST CAROLINA.  
 XX  
 PI Nyce JM;  
 DR WPI; 1999-229400/19.  
 XX  
 DR New antisense oligonucleotides used in treatment of, e.g. pulmonary  
 PT vasoconstriction  
 PT  
 XX  
 XX Disclosure; Page 37; 120pp; English.  
 XX  
 CC The specification describes antisense oligonucleotides (AAX52869-X55271)  
 CC directed against at least 2 mRNAs selected from target genes, coding and  
 CC non-coding regions of RNAs corresponding to target genes, gene  
 CC initiation codons, genomic flanking regions, intron-exon borders, the  
 CC 5'-end, the 3'-end and the juxta-section between coding and non-coding  
 CC regions and all segments of RNAs encoding proteins associated with one  
 CC or more diseases, conditions or mixtures. The antisense oligonucleotides  
 CC may be derived from sequences AAX55272-74. These multiple target  
 CC oligonucleotides (specifically AAX5180-271) can be used for the  
 CC antisense treatment of diseases and conditions. Typical diseases and  
 CC conditions are those associated with impaired respiration and

CC inflammation, including lung diseases, pulmonary vasoconstriction,  
 CC inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded  
 CC respiration, respiratory distress syndrome, pain, cystic fibrosis,  
 CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic  
 CC obstructive pulmonary disease (COPD), and cancers such as leukemias,  
 CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,  
 CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,  
 CC or have metastasized to the lungs, including breast and prostate cancer.  
 XX

Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;  
 SQ

Query Match 3.1%; Score 80.4; DB 20; Length 114955;  
 Best Local Similarity 34.0%; Pred. No. 3.1e-05;  
 Matches 159; Conservative 55; Mismatches 253; Indels 0; Gaps 0;

Qy 25 GTGGCGAGAGCCGAGCTGATGAGCGGGAGCCGACCCCTGGCCGCTGCGGGA 84  
 Db 104878 GNNHNNNSVGGCCVCGCGGNNHNNNSCVGGCCVCGGNNHNNNSCVGGCCVCGGNN 104819  
 Qy 85 CCCGAGAGACTGAGACATGAGAGCGGTGGCGGTGGCCGCGGCTGCGGCTTCTCC 144  
 Db 104818 NSCCVGGCCVCGCGGNNHNNNSGCCVGGCCVCGGNNHNNNSGCCVGGCCVCGGNN 104759  
 Qy 145 TGGCCGCGCGCGCGCGCGCGCGAGCGAGCCCGGAGCGCGCGCGCGCGCGCG 204  
 Db 104758 HNNNSGCGCCVCGCGCGGNNHNNNSCGCGCCVCGCGGNNHNNNSGCGCGCC 104699  
 Qy 205 TCGTGAGCCGCGCTGCTGAGCGAGCCCGCGGCGGCTTCTGCTGCTGAGAGCGG 264  
 Db 104698 VGGCCVCGGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104639  
 Qy 265 CTCTGAGCTGCAAGCGCGCTTGAACACTTACAGCTGGCGCGCGCGCGCGCGCG 324  
 Db 104638 CGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104579  
 Qy 325 TCGGAGTGGCGGCTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 384  
 Db 104578 NNHNNNSCG 104519  
 Qy 385 ACTTGTGAGCTGCAAGCTGAGCGGTGCTGCGCTCTACGCTGCGCGCGCGCACTGC 444  
 Db 104518 GNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGGNNHNNNSGCGCGCGCGCGCG 104459  
 Qy 445 CAGCCGTCGCGGAGAGCTGACCGAGCGCCAGCGCCACAGATACCGC 491  
 Db 104458 CVGCGGNNHNNNSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104412

RESULT 15  
 AAT91361/C  
 ID AAT91361 standard; DNA; 4020 BP.  
 XX  
 XX AAT91361;  
 AC  
 DT 14-APR-1998 (first entry)  
 XX  
 DE Orf virus genomic DNA BamHI B and BamHI G fragments.  
 XX  
 KW Parapoxvirus; vector; vaccine; antigen; HIV; herpes simplex virus;  
 KW Echinosococcus granulosis; Trichostromylylus; Haemophilus;  
 KW Osterlagia; Taenia ovis; ds.  
 XX  
 OS Orf virus strain NZ-2.  
 XX  
 FH Key location/Qualifiers  
 FT CDS complement (27..589)  
 FT /\*tag= a  
 FT /label= E1L  
 FT /note= "putative E1L coding sequence"  
 FT promoter complement (662..694)  
 FT /\*tag= b  
 FT /label= PE1L



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2004, 09:06:34 ; Search time 5297 Seconds

(without alignments)  
11814.995 Million cell updates/sec

Title: US-09-836-613-1

Perfect score: 2575  
Sequence: 1 CCCGAGGCTAGCCTTCGGGT.....AGTCAGCGCGCCGCAATTC 2575

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1:	em_estba:*
2:	em_estnum:*
3:	em_estin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_hic:*
9:	gb_est1:*
10:	gb_est2:*
11:	gb_hic:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	em_gss_hum:*
18:	em_gss_inu:*
19:	em_gss_pin:*
20:	em_gss_vit:*
21:	em_gss_vit:*
22:	em_gss_mam:*
23:	em_gss_mus:*
24:	em_gss_pro:*
25:	em_gss_rnd:*
26:	em_gss_rnd:*
27:	em_gss_vit:*
28:	gb_gss1:*
29:	gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	929.4	36.1	952	12	BM473725	AGENCOURT
2	800.2	31.1	904	13	BU182537	AGENCOURT
3	791.8	30.7	808	12	BI67002	603291867
4	789.6	30.7	925	13	BQ893829	AGENCOURT

5	753	29.2	989	12	BG760547	BG760547 602716985
6	748.6	29.1	770	12	BQ001701	BQ001701 UT-H-DH1
7	748.4	29.1	757	12	BQ014565	BQ014565 UT-H-BD1
8	745	28.9	773	12	BM976832	BM976832 UT-CF-EN1
9	743	28.9	788	9	AT118592	AT118592
10	741.8	28.8	1097	10	BG327357	BG327357
11	734.8	28.7	758	12	BM976494	BM976494 UT-CF-EN1
12	722.4	28.1	829	12	BG910492	BG910492 602806176
13	707.6	27.5	816	10	BG476070	BG476070 602520954
14	702	27.3	707	13	BQ772664	BQ772664 UT-H-PEO-
15	701.4	27.2	707	12	BQ006037	BQ006037 UT-H-EI1-
16	684.2	26.6	725	12	BM726804	BM726804 UT-E-BJO-
17	682.4	26.5	707	12	BQ108688	BQ108688 UT-H-DH1-
18	675	26.2	699	12	BQ632129	BQ632129 UT-H-FEL-
19	670.4	26.0	676	13	BU683837	BU683837 UT-CF-ENO
20	663.8	25.4	1201	13	BX417673	BX417673 BX417673
21	655.2	25.3	800	10	CA439269	CA439269 UT-H-DT1-
22	652.6	25.3	978	13	BF338953	BF338953 602036023
23	652	25.3	978	13	BU191481	BU191481
24	651.4	25.3	929	10	BG107971	BG107971 602278267
25	649.4	25.2	848	10	BG117672	BG117672 602349632
26	645.4	25.1	692	9	A1436192	A1436192
27	643	25.0	648	13	BU630517	BU630517 UT-H-FLO-
28	640.6	24.9	974	13	BQ677244	BQ677244 AGENCOURT
29	638.8	24.8	947	13	BU542714	BU542714 AGENCOURT
30	633.4	24.6	741	10	BE410798	BE410798 601301429
31	633	24.6	633	12	BM716556	BM716556 UT-E-DX1-
32	629	24.4	640	10	BF346260	BF346260 602018449
33	627.2	24.4	654	12	BM728086	BM728086 UT-E-EJO-
34	618.8	24.0	655	9	AM170220	AM170220 xm57a12.x
35	614	23.8	625	10	BG149216	BG149216 mad24507.
36	613.2	23.8	926	13	BU591032	BU591032 AGENCOURT
37	612	23.8	612	9	A1686304	A1686304
38	610.4	23.7	612	9	AM409366	AM409366
39	606.2	23.5	623	12	BM784286	BM784286 K-EST0062
40	596.2	23.2	626	10	BE219892	BE219892 hv64D03.x
41	583	22.6	583	12	BM773017	BM773017 K-EST0057
42	581.2	22.6	824	12	BI557975	BI557975 603237595
43	578.4	22.5	907	13	BX454545	BX454545 603237595
44	576	22.4	696	14	CB473703	CB473703
45	570	22.1	956	13	BQ922969	BQ922969 AGENCOURT

## ALIGNMENTS

RESULT 1  
BM473725  
LOCUS  
DEFINITION  
AGENCOURT 6484269 NIH\_MGC\_72 Homo sapiens cDNA IMAGE:5518154  
5', mRNA sequence.  
ACCESSION  
BM473725  
VERSION  
BM473725.1 GI:18522767  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 952)  
NIH-MGC <http://imgc.ncl.nih.gov/>  
AUTHORS  
National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE  
JOURNAL  
COMMENT  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-rc@mail.nih.gov](mailto:cgabs-rc@mail.nih.gov)  
Tissue Procurement: ATCC/DC/DMP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
plate: LHAM12230 row: j column: 11  
High quality sequence stop: 679.



FEATURES  
source

Location/Qualifiers  
1. .952  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5538154"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: PCMV-SPORT6; Site: 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies"

BASE COUNT 172 a 304 c 288 g 187 t 1 others  
ORIGIN

Query Match 36.1%; Score 929.4; DB 12; Length 952;  
Best Local Similarity 99.3%; Pred. No. 2.2e-164;  
Matches 933; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

950 GGAAGACCCCATATTCCTCCATCATATGAGAGGCTCTCTCTGAGAGCTGATCAAGAGTT 1009  
13 GGAAGACCCCATATTCCTCCATCATATGAGAGGCTCTCTCTGAGAGCTGATCAAGAGTT 72  
1010 TGGCAGACACACATCTATGAGGAGCGGACACTTTCATAGATGACAGCCACTTCTCAGA 1069  
73 TGGCAGACACACATCTATGAGGAGCGGACACTTTCATAGATGACAGCCACTTCTCAGA 132  
1070 GCCCTCTACCTTGGCGGAGCCACCACTGCGCTATGAGGCACTGACTGATGATAC 1129  
133 GCCCTCTACCTTGGCGGAGCCACCACTGCGCTATGAGGCACTGACTGATGATAC 192  
1130 TGAAGCTGTGTGGCTGCTCCAGAGGCTGCTCTTCACAGACAGCGGAGTTCTGGGGCC 1189  
193 TGAAGCTGTGTGGCTGCTCCAGAGGCTGCTCTTCACAGACAGCGGAGTTCTGGGGCC 252  
1190 CGCCCAATCAAGGCTGTGTGGAGGCTGTGCGCCGCTGCTCTGTTCTGAACT 1249  
253 CGCCCAATCAAGGCTGTGTGGAGGCTGTGCGCCGCTGCTCTGTTCTGAACT 312  
1250 GTTGTGAGAGGCGAGCTGTGTATACCGGCACTGCGCTCTTCAGAGGCGGAGCCCTTCA 1309  
313 GTTGTGAGAGGCGAGCTGTGTATACCGGCACTGCGCTCTTCAGAGGCGGAGCCCTTCA 372  
1310 CTGTGTGATGTGTGCAAACTTTGGGGAAACATGCTTTTGTGAGAGCCCTAAGAGCTGT 1369  
373 CTGTGTGATGTGTGCAAACTTTGGGGAAACATGCTTTTGTGAGAGCCCTAAGAGCTGT 432  
1370 GAAAGGAGGCGGAGAGGCTGCGCCGCTCTTCCCAACTCCACATGATGAGGCAAGGCGAT 1429  
433 GAAAGGAGGCGGAGAGGCTGCGCCGCTCTTCCCAACTCCACATGATGAGGCAAGGCGAT 492  
1430 GAGCCCGAGGCGATCAGCCAGAGAGAGAGGATGATCTATTCCTCACTGAGTGAAGCTG 1489  
493 GAGCCCGAGGCGATCAGCCAGAGAGAGAGGATGATCTATTCCTCACTGAGTGAAGCTG 552  
1490 GCGAAGAGACCCAGTGTCCAGATTTTGGGAGCTGCGGATGACAGCTTTGCGCCGCGGTA 1519  
553 GCGAAGAGACCCAGTGTCCAGATTTTGGGAGCTGCGGATGACAGCTTTGCGCCGCGGTA 612  
1550 TGGGCTCTCCACCGGAGCGAGGAGGAGCTGAGAGCTACTGCTCCGAGTGTGTCAA 1609  
613 TGGGCTCTCCACCGGAGCGAGGAGGAGCTGAGAGCTACTGCTCCGAGTGTGTCAA 672  
1610 CTGCTCCGAGGAGGCTGAGAGGAGCGCAATGTATGCGGCTGTCAGGCGGCGTCTCT 1669  
673 CTGCTCCGAGGAGGCTGAGAGGAGCGCAATGTATGCGGCTGTCAGGCGGCGTCTCT 732  
1670 ACAGATGAATACAGAGCTGTGTGTAACAGCATCTGATGTGTTGAGGCTGGGCGCTGCT 1729  
733 ACAGATGAATACAGAGCTGTGTGTAACAGCATCTGATGTGTTGAGGCTGGGCGCTGCT 792  
1730 GCTCACTATGTCTCCCTCCCTGCGCAACAGCCCGGCTTCGCGTACGACGCTGTGAGCT 1789

Db 793 GCTCACTATGTCTCCCTCCCTGCGCAACAGCCCGGCTTCCCTTACAGCTGTGAGCT 852  
Qy 1790 CACTGGGAGGAGTGTGAGAGAGAGCTGATGAGTGTGATGAGAGAGGAGAGGCGCTA 1849  
Db 853 CACTGGGAGGAGTGTGAGAGAGAGCTGATGAGTGTGATGAGAGAGGAGAGGCGCTA 912  
Qy 1850 CCTGAGCAAGAGAGTGTGAGGCTCCCTGTTGAGAGGCTGAGGC 1899  
Db 913 CCTGAGCAAGAGAGTGTGAGGCTCCCTGTTGAGAGGCTGAGGC 952

RESULT 2  
BUI82537  
LOCUS BUI82537 904 bp mRNA linear EST 04-SEP-2002  
DEFINITION AGENCOURT\_7946518 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6151074  
5', mRNA sequence.  
ACCESSION BUI82537  
VERSION BUI82537.1 GI:22696521  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 904)  
NIH-MGC <http://mgi.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DTF  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/BLN at:  
<http://image.llnl.gov>  
Plate: L14M13486 row: 1 column: 19  
High quality sequence stop: 598.

FEATURES  
source

Location/Qualifiers  
1. .904  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6151074"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: PCMV-SPORT6; Site: 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies"

BASE COUNT 171 a 279 c 283 g 171 t  
ORIGIN

Query Match 31.1%; Score 800.2; DB 13; Length 904;  
Best Local Similarity 96.1%; Pred. No. 3.9e-140;  
Matches 864; Conservative 0; Mismatches 28; Indels 7; Gaps 4;

Qy 1463 CTATTCCTCATGTGCTGAGCTGGCTGGCGAAGAGCCAGTGTGCAATTTGGAGGCTG 1522  
Db 1 CTATTCCTCATGTGCTGAGCTGGCTGGCGAAGAGCCAGTGTGCAATTTGGAGGCTG 60  
Qy 1523 GGTGACACACTTTGCGCGCGCGGAGTATGGGATCTCCACCGGAGCGAGGCGAGCGTG 1582  
Db 61 GGTGACACACTTTGCGCGCGCGGAGTATGGGATCTCCACCGGAGCGAGGCGAGCGTG 120  
Qy 1583 GAGGCTACTGCTCCGAGTGTGTAACTGCTCCGAGGAGGCTGTGAGAGGCGCAACATG 1642  
Db 121 GAGGCTACTGCTCCGAGTGTGTAACTGCTCCGAGGAGGCTGTGAGAGGCGCAACATG 180  
Qy 1643 TAGCCGCTGTGTGAGGCGGCGCTCCCTACAGATGAATACAGCATCTGTGTAAACGATC 1702

Db 181 TAGCCCTGCTGAGGCGGCGCTCCCTACAGATGATACAGCATCTGTGTAACAACCGATC 240  
 QY 1703 TGATGTTTGAAGGCGCTGCGGCTGCTGTCAATCTGTCTCCCTCCGCGCAACGAGCC 1762  
 Db 241 TGATGTTTGAAGGCGCTGCGGCTGCTGTCAATCTGTCTCCCTCCGCGCAACGAGCC 300  
 QY 1763 CGCTTCCGCTACACACTGCTGACCTCACTCTGCGAGGAGTGCAGAGCTGTGAGCTT 1822  
 Db 301 CGCTTCCGCTACACACTGCTGACCTCACTCTGCGAGGAGTGCAGAGCTGTGAGCTT 360  
 QY 1823 GTACTATAGAGAGGCAAGAGCGCTCACTGAGCAAGAGTGGCTCCCTGTTGAGGCG 1882  
 Db 361 GTACTATAGAGAGGCAAGAGCGCTCACTGAGCAAGAGTGGCTCCCTGTTGAGGCG 420  
 QY 1883 TGAAGGCGCTCTGCGCTTATGAGCTGCTGCCGCACTGAGAGGCTGTGCTAGTGAAG 1942  
 Db 421 TGAAGGCGCTCTGCGCTTATGAGCTGCTGCCGCACTGAGAGGCTGTGCTAGTGAAG 480  
 QY 1943 CGGCTTCTGCTGAGGAGCTGCTGAGAGAGGCGCGAGCGGCACTGAGAGGCGCA 2002  
 Db 481 CGGCTTCTGCTGAGGAGCTGCTGAGAGAGGCGCGAGCGGCACTGAGAGGCGCA 540  
 QY 2003 GCGCGATTTCTACAGAGCAAGACCGCTACAGCTGACCTTGTGCGGCGCAGAGGCA 2062  
 Db 541 GCGCGATTTCTACAGAGCAAGACCGCTACAGCTGACCTTGTGCGGCGCAGAGGCA 600  
 QY 2063 CATCTTGACTATGCTCCAAACAGAGCTGCGGCGTGTGCGCACTACTACACCTCTG 2122  
 Db 601 CATCTTGACTATGCTCCAAACAGAGCTGCGGCGTGTGCGCACTACTACACCTCTG 660  
 QY 2123 CTGCGGCGCTTTCTCTGAGGCGCTGCTTGAACAGTGTGCGCAGAGGCACTCTTCCAA 2182  
 Db 661 CTGCGGCGCTTTCTCTGAGGCGCTGCTTGAACAGTGTGCGCAGAGGCACTCTTCCAA 720  
 QY 2183 GCACCAAG-TTTGACAAAATGCTTCTCACTGAGAGAGGCTTCTTCTGAGAGAGAG 2241  
 Db 721 GCACCAAGTTTGAACAAAATGCTTCTCACTGAGAGAGGCTTCTTCTGAGAGAGAG 780  
 QY 2242 GGTATCCCGAGCGAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2300  
 Db 781 GGTATCCCGAGCGAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
 QY 2301 TATTA-CCCCGCTGAGTGGCGCGG---CTTGTGTGATGATGATGCGCAACACTGAGG 2354  
 Db 841 TATTAACCCCGAGTGGCGCGGCTCTTGGGAGAAAATTCGCGCAACACTGAGG 899  
 RESULT 3  
 B1667002 808 bp mRNA linear EST 12-SEP-2001  
 LOCUS 603291867F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:531121 5'  
 DEFINITION mRNA sequence.  
 ACCESSION B1667002  
 VERSION B1667002.1 GI:15581235  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 808)  
 NIH-MGC http://mgc.ncl.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strauberg, Ph.D.  
 Email: rgs@bbs.fda.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki, and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LIML at:  
 http://image.llnl.gov

Plate: LIM11788 row: f column: 18  
 High quality sequence stop: 807.  
 Location/Qualifiers  
 1..808  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:531121"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"  
 /clone\_id="NIH\_MGC\_96"  
 /note="Organ: brain; Vector: pBlueScript (modified  
 pBlueScript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (Gcgag  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTT-3',  
 size-selected for average insert size 2.3 kb and  
 normalized to ROT 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."  
 BASE COUNT 106 a 266 c 303 g 133 t  
 ORIGIN  
 Query Match 30.7%; Score 791.8; DB 12; Length 808;  
 Best Local Similarity 99.6%; Pred. No. 1.5e-138;  
 Matches 804; Conservative 0; Mismatches 2; Indels 1; Gaps 1.  
 87 CGAGAGACTGAGACCATGAGAGCGGTGAGCGGCGGCGGTGGGGTCTTCTCTG 146  
 Db 3 CGAGAGACTGAGACCATGAGAGCGGTGAGCGGCGGCGGTGGGGTCTTCTCTG 62  
 QY 147 GCCGGGCGCGGGGCGCGGCGAGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 206  
 Db 63 GCCGGGCGCGGGGCGCGGCGAGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 122  
 QY 207 GTGGCCCGGCTGCTGGGGCGAGCGCCGCGGCGGAGCTTCTCGTGTGCTGAGAGCGGCT 266  
 Db 123 GTGGCCCGGCTGCTGGGGCGAGCGCCGCGGCGGAGCTTCTCGTGTGCTGAGAGCGGCT 182  
 QY 267 CTGGCTGCGCAAGCGGGGCTTGAACACCTCAAGCTTGGGGCGGCGGCGGCGGCGG 326  
 Db 183 CTGGCTGCGCAAGCGGGGCTTGAACACCTCAAGCTTGGGGCGGCGGCGGCGGCGG 242  
 QY 327 CGGATGCGCGGCTCCACCGGCGGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 386  
 Db 243 CGGATGCGCGGCTCCACCGGCGGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 302  
 QY 387 TTCTGTGCTGCGCAAGCTGCGGCTGCTGAGCTGCGGCTGCGGCGGCGGCGGCGG 446  
 Db 303 TTCTGTGCTGCGCAAGCTGCGGCTGCTGAGCTGCGGCTGCGGCGGCGGCGGCGG 362  
 QY 447 GCCGTGCGGGGAGAGCTGACCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 506  
 Db 363 GCCGTGCGGGGAGAGCTGACCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 422  
 QY 507 TGCACGCAAGCTACTCTTCTGTGTGTGTGAGTGGGCGGCTGAGAGCGAGATGAC 566  
 Db 423 TGCACGCAAGCTACTCTTCTGTGTGTGTGAGTGGGCGGCTGAGAGCGAGATGAC 482  
 QY 567 TGAATGCGCGCTGANTGGATCAACTGCGCACTGCGCTGAGAGCGGCGGCGGCGG 626  
 Db 483 TGAATGCGCGCTGANTGGATCAACTGCGCACTGCGCTGAGAGCGGCGGCGGCGG 542  
 QY 627 CAGCGGAGTGAAGCTTGGGCGGCTGAGCCAGGCGAGAGATGAATGAATGATGATG 686  
 Db 543 CAGCGGAGTGAAGCTTGGGCGGCTGAGCCAGGCGAGAGATGAATGAATGATGATG 602  
 QY 687 CCGGCTTCCCTGCGGCTGCGGCGGCGGAGAGGCACTGAGAGCGGCGGCGGCGGCGG 746  
 Db 603 CCGGCTTCCCTGCGGCTGCGGCGGCGGAGAGGCACTGAGAGCGGCGGCGGCGGCGG 661  
 QY 747 CCTCTTGGCAATCAAGAGCTTAACTGCGAGCAGCGGCTCTGAGCAAGATGCGGCTCC 806

Db	662	CCCTCTGGCAGCATCAAGAGCGTTTACCTGACGACACCGGGTCTCTGGACCAAGTGGCGTCC	721
Qy	807	TTGGGCAATGACCCCAAGTGTGCTGCTGATTCGGCGGCATGTTCCCGAGGCTGTCAACCAAG	866
Db	722	TTGGCAATGACCCCAAGTGTGCTGCTGATTCGGCGGCATGTTCCCGAGGCTGTGTCAACCAAG	781
Qy	867	GTGTTCCCTCAGTCAATGTCAACGAAG	893
Db	782	GTGTTCCCTCAGTCAATGTCCCGAAG	808
RESULT 4			
LOCUS	BQ893829	925 bp	mRNA
DEFINITION	AGENCOURT 8728368 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6339115	linear	EST 16-AUG-2002
ACCESSION	BQ893829		
VERSION	BQ893829.1	GI:22285843	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a> Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed By: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: L10CM2533 row: g column: 20 High quality sequence stop: 668. Location/Qualifiers 1..925 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6339115" /tissue_type="neuroblastoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_47" /note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size selected >50bp for average insert size 1.8kb. Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."		
BASE COUNT			
ORIGIN	164 a	290 c	292 g 179 t
Query Match			
Best Local Similarity	30.7%;	Score 789.6;	DB 13; Length 925;
Matches	893; Conservative	0; Mismatches	9; Indels 58; Gaps 4;
Qy	1114	TGACTGCGAGTGTGATGAGCTGTGTGAGCTGTGCTCAAGGCTGTGCTCTTCACGACCAAGC	1173
Db	13	TGTACACAGTGTGATGAGCTGTGTGAGCTGTGCTCTTCACGAGGCTGTGCTCTTCACGACCAAGC	72
Qy	1174	CGCAGTTCTGGGGGCGCCGCCAGATCAAGAGGCTGTGTCTGGAGCTGTGCCCCGTCGCGCC	1233
Db	73	CGCAGTTCTGGGGGCGCCGCCAGATCAAGAGGCTGTGTCTGGAGCTGTGCCCCGTCGCGCC	132
Qy	1234	TCCGAGTTCTGGAGCTGTGTGCTAGAGCCAGCCTGTGTATACCCGCACTGCTCTCTCC	1293

D	b		133	TCTGGTGTCTGAGACCTGTTTGTCGTAGAGCCAGCGCTGTATACCAGCACTGCCTCCTCC	192
O	y		1294	AGGGCCAAAGCCCTTATCTGTGTGCATGTCTGCCAATTGTTGGGGAAAACAATGGTCTTTTGG	1357
D	b		193	AGGGCAAGCCCTTATCTGTGTGCATGTCTGCCAACTTGTGGGGAAAAACATGATGTCTTTTGG	252
O	y		1354	GAGCCCTTAGAGCTGTGAACGGAAGCCCCAGAAGCTGCCCTCTTCCCACATCTCACCA	1413
D	b		253	GAGCCCTTAGAGCTGTGAACGGAAGCCCCAGAAGCTGCCCTCTTCCCACATCTCACCA	312
O	y		1414	TGTATAGCATCGGGCATGGCCCCCGAAGGCATACGACAGAACGAATGTGTATTTCCCTCA	1473
D	b		313	TGTATAGCATCGGGCATGGCCCCCGAAGGCATACGACAGAACGAATGTGTATTTCCCTCA	372
O	y		1474	TGGCTGAGCTGGGGTGGGAAAAGAACCCAGTGCAGATTGTCAGCCCTGGGTCACAGCT	1533
D	b		373	TGGCTGAGCTGGGGTGGGAAAAGAACCCAGTGCAGATTGTCAGCCCTGGGTCACAGCT	432
O	y		1534	TTGCCGCCCGCGGGTATGGGGTCTTCCACCCGACGACGAGGGCAGCGTGAAGCTACTGC	1593
D	b		433	TTGCCGCCCGCGGGTATGGGGTCTTCCACCCGACGACGAGGGCAGCGTGAAGCTACTGC	492
O	y		1594	TCCGAGATGTATATACTGTCTCCGGGAGAGCTTCAGAGGGCCAATGTATGCCCGCTGG	1653
D	b		493	TCCGAGATGTATATACTGTCTCCGGGAGAGCTTCAGAGGGCCAATGTATGCCCGCTGG	552
O	y		1654	TCAGGCGGCGCTCCCTCATAGATGAATACAGCATCTGGTAACAACGATCTGATGTATTG	1713
D	b		553	TCAGGCGGCGCTCCCTCATAGATGAATACAGCATCTGGTAACAACGATCTGATGTATTG	612
O	y		1714	AGGCTCTGGCGGCTGTCTGTCTCATATCTGTCTCTCTCTGGCCACAGGCCCGCTTCGCT	1773
D	b		613	AGGCTCTGGCGGCTGTCTGTCTCATATCTGTCTCTCTCTGGCCACAGGCCCGCTTCGCT	672
O	y		1774	ACGACCTGTCTGGA-CTCTCATCTGGCAGGAGGTGACAGGAGCTGTGTACTATATGAG	1832
D	b		673	ACGACCTGTGTGAACTCTCATCTGGCAGGAGGTGACAGGAGCTGTGTACTATATGAG	732
O	y		1833	GAGGCAAGAAAGCGCTTACTGTAGAGCAAGAGCTGGCTCTCTGTGTAGAGGCTGAGAGCGTC	1892
D	b		733	GAGGCAAGAAAGCGCTTACTGTAGAGCAAGAGCTGGCTCTCTGTGTAGAGGCTGAGAGCGTC	792
O	y		1893	CTGGCCTATGAGCTGTGCGCGGACCTGACGAGSTGTGGCTATGTAGACGCCGCTTCTTG	1953
D	b		793	CTGGCCTATGAGCT-----	806
O	y		1953	CTGGGAGACTGTGCTTAGAGACGCCCGGACAGCGCAGTCACTAGTAG--CCGAGGCCGATT	2011
D	b		807	-----TGCTTAGAGCAGGCGCCGAGCAAACGAACTGAGGCGCCGAGCGATT	857
O	y		2012	CTATAGACAGAAACAGCGGCTTACCAAGCTGACTTGT-GGGGGCCAGAAAGGCAATCTCTGG	2070
D	b		858	CTATAGAAACAGAAACAGCGGCTTACCAAGCTGAACTTGTGGGGGGGCCAGAAAGGCAATCTCTGG	917

RESULT 5  
BG760547

LOCUS  
DEFINITION  
602716985F1 NIH\_MGC\_49 Homo sapiens cDNA clone IMAGE:4840436 5',  
mRNA sequence.  
Bg760547

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
EST.

ORGANISM  
*Homo sapiens* (human)  
*Homo sapiens*  
Bokaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 989)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: rga@ds-femail.nih.gov



vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is AGATCATTTGC.

```
TAG_LIB=UI-H-DHI
TAG_TISSUE=lung
```

29

	a	c	g	t.	
BASE COUNT	146	234	215	172	3 others
ORIGIN					

VERSION	BO014565.1	GI:19739466
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 757)	
AUTHORS	NCI-CCAP	<a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a>
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	

Query Match	29.1%	Score 748.6;	DB 12;	Length 770;
Best Local Similarity	99.1%;	Pred. No. 1.9e-130;		
Matches 763; Conservative	0;	Mismatches 5;	Indels 2;	Gaps 1;

**Source**

POLYA=Yes.

1.75

```
/organism="Homo sapiens"
```

```
/mol_type="tRNA"
```

```
/usr_x11=/usr_x11:2000
/cjone="IMAGE:5833487"
```

```
/tissue type="Chondrosarcoma"
```

```
/dev_stage="Adult"
```

```
/lab_host="DH10B (L
```

```
/clone_lib="NCI_CGAP_ED1"
```

```

/note="Organ: Left Pubic Bone; Vector: pr/13-Pac
(Dbs:main) with a modified polymer: Site 1:

```

site 2: Not T: NCT CGAP ED1 is a normalized cdna (Pindlmacia) with a modified polyinker, see 1:

containing the following tissue(s): Chondrosarcoma cell

line CS5. The 1

Lennon and Soares, Geno

First strand cDNA s

QY	2148	GTTGCACTGTGGCCACAGGCACTCCCTTCCACAGCAACAGTTGACAAAATCTCTC	2207
Db	4:10	GTTGCACTGTGGCCACAGGCACTCCCTTCCACAGCAACAGTTGACAAAATCTCTC	351
QY	2208	CACTGAGCAGGCGCTTGTTTCAGCAGCAGAGGTACCCAGCAGCGCGAGAGAC	2267
Db	350	CACTGAGCAGGCGCTTGTTTCAGCAGCAGAGGTACCCAGCAGCGCGAGAGAC	291

QY 2258 ACTGTGACCTGGCCAGAGATCTTCTCAATATTAACCCGGCGTGGGTGGCCGGCTCT 232

Dp 230 ACTGTGACCTGGCCAGAGATCTTCTCAATATTAACCCGGCGTGGGTGGCCGGCTCT 231

QY 2328 TGGTGATAGATTGCGCACCACTGGGCGCTGTTTTCGCTAATTCAGGGCAGATTCCAG 2387

Db 230 TGGTGATAGATTGCGCACCACTGGGCGCTGTTTTCGCTAATTCAGGGCAGATTCCAG 171

Oy 2388 GCCCAGACTGGACAGACATCAAGAGTAAACCAGGCTGGGAGGAGGCCCCAGGCTG 2447  
 Db 170 GCCCAGAGCTGGACAGATCAAGAGTAAACCAGGCTGGGAGGAGGCCCCAGGCTG 111

Oy 2448 CTGCTGGGCTCTGAACCTGGGGGAAATTGGAGGGAAATGAACCTGCCTCCACACACACCC-- 2505  
 Db 110 CTGCTGGGCTCTGAACCTGGGGGAAATTGGAGGGAAATGAACCTGCCTCCACACACACACCCNN 51

Qy	2506	AAAGTGGGATTAAAGTACTGTTTTCTTTCCACTTAAAAA	2555
Db	50	AAAGTGGGATTAAAGTACTGTTTTCTTCCAAAAA	1

RESULT 7				
B0014565/c				
LOCUS	B0014565	757 bp	mRNA	linear
DEFINITION	UI-H-ED1-act-K-24-0-UI-s1 NCI CGAP_Ed1 Homo sapiens cDNA clone			EST 26-MAR-2007
IMAGE:5833487	3'			mRNA sequence.
ACCESSION	B0014565			

BASE COUNT	146 a	230 c	207 g	170 t	4 others
ORIGIN					

Query Match	29.1%	Score	748.4;	DB	12;	Length	757;
Best Local Similarity	99.3%	Pred.	No. 2.4e-130;				

Matches	749,	Conservative	0,	Mismatches	5,	Indels	0,	Gaps	0;
QY	1803	GTGCAGAGCTGTCAGCTTGTACTATCAGAGCGAAGAACGCGCTACTAGCAGAGAG	1862						

DB	757	GTGGAGGAGCTGCTCAGATTGTACTATGAGGAGGCAAGACGCGTACNTAGGCAAGAG	698
QY	1863	CTGCGCTCCCTGTGAGGAGCTGAGGAGCTCTGGGCTATGATGACTGCTGCCGGA	1922

Db 697 CTGCCCCCCTGTTGAGGGGCTGGAGAGCGTCTTGCGCTATGAGCGTCGTCGCCACACTGGAC 638

Qy 1923 GAGGTCGTGCTAGTGACAGCGCGTCTTTCGTGGAGCACTGGCTAAGACAGAGCCCGAGCA 1982

Db 637 GAGGAGCTGGCTAGTGAACAGCCGCTCTTGTCTGGCAGCTGGCTATGAGACAGGCCCGGAGACA 5/8

Qy 1983 GCGGCACTCACTGAGGCCGAGGCCGATTTCTACAGACGAAACAGCCGCTACCAAGTAAAC 2042

D6	577	GCGGCAGTCAGTGAGGCCGAGGCCGATTATTACAGACGACCAACCCGATACACAGTGCAC	318
Qy	2043	TTGTGTGGGGGCGCAGAGGCGCATCTCTGGACTATGCCAACAAGCTGTGGCGGGTGTGTG	2102

Db 517 TTGTGGGAGCCAGAGGCAACATCTGAGATATGCCAAGACAGCTGGCGGGTTGTTG 458

QY 2103 GCCAATCTACTACACCCCTCGCTGGGGCTTTTCTGAGAGCGCTGCTGACAGTGGGCC 2162

Db 457 GCCAATCTACTACACCCCTCGCTGGGGCTTTTCTGAGAGCGCTGCTGACAGTGGGCC 398

QY 2163 CAGGAGCATCCCTTTCCAAACAGACACCAATTGACAAAATATGCTTCCAACTGAGCAGGCC 2222

Db 397 CAGGAGCATCCCTTTCCAAACAGACACCAATTGACAAAATATGCTTCCAACTGAGCAGGCC 338

QY 2223 TTGCTTTCTCAGACAGAGGTACCCAGCCAGCCGAGAGACACTGTGAGCTGGGCC 2282

Db 337 TTGCTTTCTCAGACAGAGGTACCCAGCCAGCCGAGAGACACTGTGAGCTGGGCC 278

QY 2283 AAGAAGATCTTCTCAATATATTACCCCGCTGGCTGGCCGCTTGTGTATGATTGCG 2342

Db 277 AAGAAGATCTTCTCAATATATTACCCCGCTGGCTGGCCGCTTGTGTATGATTGCG 218

QY 2343 CACCACTGGGCTGCTTTTCCGCTAATTCCAGGCAATTCAGGGCCCAAGCTGGACA 2402

Db 217 CACCACTGGGCTGCTTTTCCGCTAATTCCAGGCAATTCAGGGCCCAAGCTGGACA 158

QY 2403 GACATCACAGATTAACCCAGGCTGGAGAGAGCCCAAGGCTGCTGTGGGGTTGAC 2462

Db 157 GACATCACAGATTAACCCAGGCTGGAGAGAGCCCAAGGCTGCTGTGGGGTTGAC 98

QY 2463 CTGGGGGGATTGAGAGGAATGACCTGCTCCACCAACCAAGTGTGGATTAAAG 2522

Db 97 CTGGGGGGATTGAGAGGAATGACCTGCTCCACCAACCAAGTGTGGATTAAAG 38

QY 2523 TACTGTTTCTTCCACTTAAAAAATAAAAA 2556

Db 37 TACTGTTTCTTCCACTTAAAAAATAAAAA 4

RESULT 8  
BM976832/c 773 bp mRNA linear EST 21-FEB-2003

LOCUS UI-CF-EN1-adb-o-24-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone

DEFINITION UI-CF-EN1-adb-o-24-0-UI.3', mRNA sequence.

ACCESSION BM976832

VERSION BM976832.1 GI:19594633

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 773)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 8889548

COMMENT Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com) or from Open Biosystems (www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA=yes.  
location/Qualifiers  
1..773  
/organism="Homo sapiens"

/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-EN1-adb-o-24-0-UI"  
/tissue\_type="Primary Lung Cystic Fibrosis Epithelial Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-EN1"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to a EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.  
TAG\_LIB=UI-CF-EN1  
TAG\_TISSUE=Human Lung Epithelial Cell Lines untreated LPS 6hr to LPS 24h  
TAG\_SEQ=CTGCTCAGGT"

BASE COUNT 147 a 232 c 215 g 174 t 5 others

ORIGIN

Query Match 28.9%; Score 745; DB 12; Length 773;  
Best Local Similarity 98.6%; Pred. No. 9.1e-130;  
Matches 759; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1788 CTCATCGCAGCAGTGCAGAGCTGTCAGCTTGTACTATGAGAGGCAAGAGCCG 1847

Db 773 CTCATCGCAGCAGTGCAGAGCTGTCAGCTTGTACTATGAGAGGCAAGAGCCG 714

QY 1848 TACCTGAGAAAGACCTGCTCCCTGTTGAGGGCTGAGAGGCTGCTGATAGAGTG 1907

Db 713 TACCTGAGAAAGACCTGCTCCCTGTTGAGGGCTGAGAGGCTGCTGATAGAGTG 654

QY 1908 CTGCCGCACTGAGAGAGTGTGCTAGTGAACCGCTTCTGCTGGAGCTGAGCTA 1967

Db 653 CTGCCGCACTGAGAGAGTGTGCTAGTGAACCGCTTCTGCTGGAGCTGAGCTA 594

QY 1968 GAGCAGGCCCGAGCGCGCACTGAGAGCCGAGCCGATTTCTAGAGGAAACAGC 2027

Db 593 GAGCAGGCCCGAGCGCGCACTGAGAGCCGAGCCGATTTCTAGAGGAAACAGC 534

QY 2028 CGCTACAGCTACCTTGTGGGGGCGAGAGGCAATCTGACTATGCCAAGAGC 2087

Db 533 CGCTACAGCTACCTTGTGGGGGCGAGAGGCAATCTGACTATGCCAAGAGC 474

QY 2088 CTGGCGGGGTGTGTGCCAATCTAACAACCCCTGCTGGCGGCTTTCTGAGAGCGCTG 2147

Db 473 CTGGCGGGGTGTGTGCCAATCTAACAACCCCTGCTGGCGGCTTTCTGAGAGCGCTG 414

QY 2148 GTTGACAGTGTGGCCAGAGGCACTCTTCCAAAGACCAAGTTGACAAAATGTCTTC 2207

Db 413 GTTGACAGTGTGGCCAGAGGCACTCTTCCAAAGACCAAGTTGACAAAATGTCTTC 354

QY 2208 CAATGAGACAGCCCTTGTCTGACAGAGAGTACCCAGCCAGCCGAGAGAGC 2267

Db 353 CAATGAGACAGCCCTTGTCTGACAGAGAGTACCCAGCCAGCCGAGAGAGC 294

QY 2268 ACTGTGAGCTGGGCAAGAAATCTTCTCAAAATTTCACCGGCTGGGTGGCGGCTCT 2327

Db 293 ACTGTGAGCTGGGCAAGAAATCTTCTCAAAATTTCACCGGCTGGGTGGCGGCTCT 234

QY 2328 TGGTATGATTCGCACCACTGGGCTTGTTCGGTAATTCAGGGCAGATTCCAGG 2387

Db 233 TGGTATGATTCGCACCACTGGGCTTGTTCGGTAATTCAGGGCAGATTCCAGG 174

```

QY 2388 GCCCAGAGCTGACAGACATCACAGGATPACCAGCCTCGAGAGAGGCCCCACGGCCTG 2447
Db 173 GCCCAGAGCTGACAGACATCACAGGATPACCAGCCTCGAGAGAGGCCCCACGGCCTG 114
QY 2448 CTGGTGGGGCTGACCTGGGGGGGATTGGAGGGGAATGACTGGCCTCCAGCACCAGCC-A 2506
Db 113 CTGGTGGGGCTGACCTGGGGGGGATTGGAGGGGAATGACTGGCCTCCAGCACCAGCCAA 54
QY 2507 AAGTGGGATTAAGTACTGTTTCTTCCACTTAAAAA 2556
Db 53 AAGTGGGATTAAGTACTGTTTCTTCCACTTAAAAA 4

RESULT 9
AUI18592 788 bp mRNA linear EST 01-AUG-2002
LOCUS AUI18592 HEMBA1 Homo sapiens CDNA clone HEMBA1003940 5', mRNA
DEFINITION
ACCESSION AUI18592
VERSION AUI18592
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
Yamanoto, J., Nakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
Isogai, T.
TITLE HRI human cDNA project
JOURNAL Unpublished
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kasarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute
Location/Qualifiers
1. 788
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEMBA1003940"
/cisue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/clone_1ib="HEMBA1"
/notes="Vector: PME18SFL3"
BASE COUNT 135 a 250 c 236 g 164 t 3 others
ORIGIN
Query Match 28.9%; Score 743; DB 9; Length 788;
Best Local Similarity 98.3%; Pred. No. 2.2e-129;
Matches 760; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 1014 ACAGACCACTATAGGGGCGGACACTTTCATGAGATGAGGACGACCTTCTCAGAGCCC 1073
Db 1 ACAGACCACTATAGGGGCGGACACTTTCATGAGATGAGGACGACCTTCTCAGAGCCC 60
QY 1074 TCTCACTTGGCCGACGACCACTGGCGGTATAGAGCCATGATGATGATGATGAG 1133
Db 61 TCTCACTTGGCCGACGACCACTGGCGGTATAGAGCCATGATGATGATGATGAG 120
QY 1134 GCTGTGTGCTGCTCCAGAGCTGGCTCTTCCAGACACGCGGAGTTCTGGGGGCCGCC 1193
Db 121 GCTGTGTGCTGCTCCAGAGCTGGCTCTTCCAGACACGCGGAGTTCTGGGGGCCGCC 180
QY 1194 CAGATCAGGGCTGTGCTGGAGAGCTGTGCCCGCGGCTCTGTGTTCTGACCTGTTT 1253

```

```

Db 181 CAGATCAGGGCTGTGCTGGAGCTGTGCCCCGTGGCCGCTCTCTGATCTGACCTGTT 240
QY 1254 GCTGAGAGCCAGCTGTGTATATACCCGACTGCTCTTTCAGAGGCCAGCTTTCATCTGG 1313
Db 241 GCTGAGAGCCAGCTGTGTATATACCCGACTGCTCTTTCAGAGGCCAGCTTTCATCTGG 300
QY 1314 TGCAATGTCACACCTTGTGGGGGAACCATGAGCTTTTGAAGCCCTAGAGGCTGTGAAC 1373
Db 301 TGCAATGTCACACCTTGTGGGGGAACCATGAGCTTTTGAAGCCCTAGAGGCTGTGAAC 360
QY 1374 GAGAGCCAGAGAGCTGCCGCTCTTCCCACTCCACATGTGTAGGCAGGAGCATGGCC 1433
Db 361 GAGAGCCAGAGAGCTGCCGCTCTTCCCACTCCACATGTGTAGGCAGGAGCATGGCC 420
QY 1434 CCCGAGGGCATGACGACAGAACAGAGTGTATTTCCCTATGAGCTAGCTGGAGCTGGCA 1493
Db 421 CCCGAGGGCATGACGACAGAACAGAGTGTATTTCCCTATGAGCTAGCTGGAGCTGGCA 480
QY 1494 AAGGACCCAGTGCAGATTTGGCAGGCTGGGTACAGCTTTGGCGCCCGCGATGGG 1553
Db 481 AAGGACCCAGTGCAGATTTGGCAGGCTGGGTACAGCTTTGGCGCCCGCGATGGG 540
QY 1554 GTCTCCCACTCCGAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1613
Db 541 GTCTCCCACTCCGAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
QY 1614 TCCGAGGAGGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1673
Db 601 TCCGAGGAGGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
QY 1674 ATGATATCCAGCATCTGTGTACAA-CCGATCTGATGTGTTTGAAGCCCTGGAGGCTGTGCT 1732
Db 661 ATGATATCCAGCATCTGTGTACAA-CCGATCTGATGTGTTTGAAGCCCTGGAGGCTGTGCT 720
QY 1733 CACATCTGCTCCTCCTCGGACCAAGCCCGGCTTCCGCTACAGACTGTGG 1785
Db 721 CACATCTGCTCCTCCTCGGACCAAGCCCGGCTTCCGCTACAGACTGTGG 773

RESULT 10
B327597 1097 bp mRNA linear EST 27-FEB-2001
LOCUS B327597 NIH_MGC_14 Homo sapiens CDNA clone IMAGE:4564388 5',
DEFINITION mRNA sequence.
ACCESSION B327597
VERSION B327597.1 GI:13134035
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE NIH-MGC Http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: csapbe-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA sequencing: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLT at:
http://image.llnl.gov
Plate: L1CM1279 row: 1 column: 21
High quality sequence stop: 783.
Location/Qualifiers
1. 1097
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4564388"
/tissue_type="renal cell adenocarcinoma"

```



```

/lab host="DH10B (phage-resistant)"
/clone.lib="NH10B"
/notes="Organ: Kidney; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCAAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 212 a 340 c 338 g 205 t 2 others

Query Match 28.8%; Score 741.8; DB 10; Length 1097;  
Best Local Similarity 91.5%; Pred. No. 3.5e-129;  
Matches 843; Conservative 0; Mismatches 67; Indels 11; Gaps 5;

QY 967 CCATCATCGGAGGCTCTTCTCTGAGAGCTGATCAAAGATTGGACACACATCT 1026  
DB 2 CCATCATCGGAGGCTCTTCTCTGAGAGCTGATCAAAGATTGGACACACATCT 61  
QY 1027 ATGGAGGCGGACACTTTCATAGATGAGGACCACTTCTCAGAGCCCTCTACCTTGGCG 1086  
DB 62 ATGGAGGCGGACACTTTCATAGATGAGGACCACTTCTCAGAGCCCTCTACCTTGGCG 121  
QY 1087 CAGCCACACTGCGCTGATAGAGGCAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1146  
DB 122 CAGCCACACTGCGCTGATAGAGGCAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 181  
QY 1147 TCAGAGGCTGAGCTTCCAGACACGAGTTCAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1206  
DB 182 TCAGAGGCTGAGCTTCCAGACACGAGTTCAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 241  
QY 1207 TGGTGGAGGCTGAGCTTCCAGACACGAGTTCAGAGGCGCGCGCGCGCGCGCGCGCGCGCG 1266  
DB 242 TGGTGGAGGCTGAGCTTCCAGACACGAGTTCAGAGGCGCGCGCGCGCGCGCGCGCGCGCG 301  
QY 1267 CTGTGTATACCCGCACTGCTCTCTTCCAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1326  
DB 302 CTGTGTATACCCGCACTGCTCTCTTCCAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 361  
QY 1327 ACTTTGGGGGAAACATGATGTTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 1386  
DB 362 ACTTTGGGGGAAACATGATGTTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 421  
QY 1387 CTGCGCGCTCTTCTCCCACTTCCACATGAGTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 1446  
DB 422 CTGCGCGCTCTTCTCCCACTTCCACATGAGTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 481  
QY 1447 GCGAGAAAGAGGCTATTTCCCTCAATGAGTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 1506  
DB 482 GCGAGAAAGAGGCTATTTCCCTCAATGAGTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 541  
QY 1507 CAGATTGGAGGCTTGGTGAAGCTTGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1566  
DB 542 CAGATTGGAGGCTTGGTGAAGCTTGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 601  
QY 1567 ACAGAGGCGGAGCTGAGAGCTTCTGCTCGAGAGTGTGTAACAATGCTCCGGGAGAGGCT 1626  
DB 602 ACAGAGGCGGAGCTGAGAGCTTCTGCTCGAGAGTGTGTAACAATGCTCCGGGAGAGGCT 661  
QY 1627 GCAAGGGGCAAAATGAGCGCGTGTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1686  
DB 662 GCAAGGGGCAAAATGAGCGCGTGTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
QY 1687 TCTGTGTCAACCATGATGTGTGTTGAGGCGCTGGCGCGCTGCTCAATCTGCTCCCT 1746  
DB 721 TCTGTGTCAACCATGATGTGTGTTGAGGCGCTGGCGCGCTGCTCAATCTGCTCCCT 779  
QY 1747 CCTGTGCAACAGCCCGCTTCCGCTACACCTGTGAGCTTCACTGTGGCAGGAGTGC 1806  
DB 780 GGC-----ACAAACCGCGCTTGGGTA-GAAGTGTGAGCTCACTGGC--GAGAGTC 830

QY 1807 AGAGCTGTGAGCTTGTACTATGAGAGGCAAGAGCGCTTACCTGAGCAGAGCTGG 1866  
DB 831 AGAAGCGTACCTGTGAGCAATGAGAGAGGAGCAAAAGGCGCAACCTGTGAGAGGCGCGCG 890  
QY 1867 CCTGCTGTGAGGCGCTGAG 1887  
DB 891 CCGCGCTTGAAGGCGCGCGCAG 911

## RESULT 11

BM976494/c

LOCUS

DEFINITION

BM976494

UI-CF-EN1-acz-1-12-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone

UI-CF-EN1-acz-1-12-0-UI 3', mRNA sequence.

BM976494

## ACCESSION

BM976494.1

GI:19594088

EST

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 758)

## KEYWORDS

Normalisation and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

8889548

McCrack Lab

University of Iowa

## JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: McCray, PB

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

## REFERENCE

AUTHORS

TITLE

Normalisation and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

8889548

## JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: McCray, PB

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

## REFERENCE

AUTHORS

TITLE

Normalisation and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

8889548

## JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: McCray, PB

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

## REFERENCE

AUTHORS

TITLE

Normalisation and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

8889548

## JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: McCray, PB

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

## REFERENCE

AUTHORS

TITLE

Normalisation and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

8889548

## JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: McCray, PB

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

## REFERENCE

AUTHORS

TITLE

Normalisation and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

8889548

## JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: McCray, PB

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

## REFERENCE

AUTHORS

TITLE

Normalisation and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

8889548

## JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: McCray, PB

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

## REFERENCE

AUTHORS

TITLE

Normalisation and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

8889548

## JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: McCray, PB

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

## REFERENCE

AUTHORS

TITLE

Normalisation and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

8889548

## JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: McCray, PB

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

## REFERENCE

AUTHORS

TITLE

Normalisation and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

8889548



D <b>b</b>	538	TTGGAGGCGCACTTAAACGTGTCCATCCTCGTCCTTCTCTTGAGCTCCGGAAGAACC	597
O <b>y</b>	962	ATTCCCCATCATCGGAGAGCCTTCTCTCGAGAGCTGATCAAAGATTGGCACAGACCA	1021
D <b>b</b>	598	ATTCCCACATCATCGGAGGCCTTCTCTCGAGAGCTGATCAAAGATTGGCACAGACCA	657
O <b>y</b>	1022	CATCATGAGGAGCCGACACTTTCATGAGATGAGACCACTTCTCTGAGAGCCTCTTACT	1081
D <b>b</b>	658	CATCATGAGGAGCCGACACTTTCATGAGATGAGACCACTTCTCTGAGAGCCTCTTACT	717
O <b>y</b>	1082	TGCCGACGCCACCACTGCGCTGTATGAGGCCATGATG-CAGTGSANACTAAGC--TGT	1138
D <b>b</b>	718	TGCCGACGCCACCAATGCGCTGTATGAGGCCATGATGCGCATGATCTCAGGCGCTTCGT	777
O <b>y</b>	1139	GTCGCTGCTGC--AAGGCTGGCTCTTCCAGACCCGAGCGCAGTTTGAGGGG	1188
D <b>b</b>	778	CTGGCTGCTCCAAAGGCGCTGGCGCTTTCACAAGNCACACCCGCGCTGTGAGAGC	829

RESULT	13
BG476070	
LOCUS	816 bp
DEFINITION	mRNA sequence.
ACCESSION	BG476070
VERSION	BG476070.1
KEYWORDS	EST.
SOURCE	Homo sapiens
ORGANISM	Homo sapiens (human)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 816)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC).  
Unpublished  
Contact: Robert Stransberg, Ph.D.

Email: csgabs-remail.nih.gov  
Tissue Procurement: ARCC/BCCT/DRP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LINC1403 row: k column: 13  
High quality sequence stop: 778.

FEATURES  
Source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon.9606"
/clone="IMAGS:4639620"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_id="NH MGC 20"
/note="Organ: skin; Vector: pOT8; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT
162 a 241 c 256 g 157 t
ORIGIN

```

Query Match	27.5%	Score	707.6	DB	10	Length	816
Best Local Similarity	97.3%	Pred	No. 9	5e-123			
Matches	794	Conservative	0	Mismatches	14	Indels	8
Gaps							7

  

QY	1682	CAGCATCTGTGTACAAACCGATCTGTGATGTGTTTGAAGCGCTTGCGCGCTCTCTCTCAACATCTTCG	1744
DB	2	CAGCATCTGTGTACAAACCGATCTGTGATGTGTTTGAAGCGCTTGCGCGCTCTCTCTCAACATCTTCG	61

QY	1742	TCCTCCCTGGACCAAGACCCCGCTTCGCGTAAAGACTCTGCGACCTCACTCGGACAGGC	1801
Db	62	TCCCTCCTCGGCCACAGCCCGCTTCGCGTAAAGACTCTGCGACCTCACTCGGACAGGC	121
QY	1802	AGTGAAGAGCTGTGACTGCTGTATCTATGAGAGAGCAAGACCGCTTACTCTGAGCAAGGA	1861
Db	122	AGTCAAGAGAGCTGTGACTGCTGTATCTATGAGAGAGCAAGACCGCTTACTCTGAGCAAGGA	181
QY	1862	GCTGAGCCCTCCCTGTATGAGAGCGCTGAGAGCGTCTGGCTTATAGCTGTGCGCGACTGGA	1921
Db	182	GCTGAGCCCTCCCTGTATGAGAGCGCTGAGAGCGTCTGGCTTATAGCTGTGCGCGACTGGA	240
QY	1922	CGAGGTGTGAGCTATGTAACAACCGCTTCTTGCTGGGCAAGCTGTAGAGCAGCTCCGAGC	1981
Db	241	CGAGGTGTGAGCTATGTAACAACCGCTTCTTGCTGGGCAAGCTGTAGAGCAGCTCCGAGC	299
QY	1982	AGCGGCACTCATGTAAGGCGGAGCGCGATTTCTAGACAGAAACAGCGCTACAGCTGAC	2041
Db	300	AGCGGCACTCATGTAAGGCGGAGCGCGATTTCTAGACAGAAACAGCGCTTACAGCTGAC	359
QY	2042	CTTGTGGGGGCGAAGAGGCAACATCTGGACTATGCAACAAGCAGCTGCGGGGTGGT	2101
Db	360	CTTGTGGGGGCGAAGAGGCAACATCTGGACTATGCAACAAGCAGCTGCGGGGTGGT	419
QY	2102	GGCCCACTACACACCCCTGCTGGCGGCTTTCTGGAAGCGTGGTGTACAGTGTGGC	2161
Db	420	GGCCCACTACACACCCCTGCTGGCGGCTTTCTGGAAGCGTGGTGTACAGTGTGGC	479
QY	2162	CCAGGGCATCCCTTTCACAACAGCAACAGATTGACAAATATGTCTTCAACTGAGCAAGC	2221
Db	480	CCAGGGCATCCCTTTCACAACAGCAACAGATTGACAAATATGTCTTCAACTGAGCAAGC	539
QY	2222	CTTGCTTCTCCAGAGAGAGAGGTACCCACGACACCGGAGGAGACACTGTGACCTGGC	2281
Db	540	CTTGCTTCTCCAGAGAGAGAGGTACCCACGACACCGGAGGAGACACTGTGACCTGGC	599
QY	2282	CAAGAAGATCTTCTCAATATATACCCCGCTGGGTGGCGCGCTCTTGGTATAG-ATTG	2340
Db	600	CAAGAAGATCTTCTCAATATATACCCCGCTGGGTGGCGCGCTCTTGGTATAGATTG	659
QY	2341	GCGACCACTGGGCTTGTTTCCGCTTATTCACAGGCGAGATTGCA-GGGGCCAAGAGCTGG	2399
Db	660	GCGACCACTGGGCTTGTTTCCGCTTATTCACAGGCGAGATTGCAAGGAGGCCCAAGAGCTGG	719
QY	2400	ACAGAT-CATCAACAGATATACCCAGCGCTGGAGAGAGGCCCAACAGCGCTGTGTGGGCTC	2458
Db	720	ACAGATCATTTACAGAGATACCCAGAGCGCTGGAGAGAGGCCCAACAGG-CTGTGTGTGGGCTC	778
QY	2459	TGACTTGGGGGAGATTGAGGGAAATGACATGCTCCCTC 2494	
Db	779	TGACTT-GGGGATTTGGGGGAAATGACTGCTTAC 812	

[illegible]





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2004, 09:08:09 ; Search time 150 Seconds  
(without alignments)  
7577.077 Million cell updates/sec

Title: US-09-836-613-1  
Perfect score: 2575  
Sequence: 1 CCGGGCTTACCTTCGGGT.....AGTCAGCGCGCGCAATTC 2575

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapexc 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2575	100.0	2575 3	US-09-077-354B-1
2	1426.6	55.4	10380 3	US-09-077-354B-3
3	70.4	2.7	5173 1	US-08-242-677-1
4	66.2	2.6	1288 1	US-08-440-856A-9
5	66.2	2.6	152331 3	US-09-128-155-16
6	62.2	2.4	4411529 3	US-08-103-840A-1
7	61	2.4	4403765 3	US-09-103-840A-2
8	60.2	2.3	4403765 3	US-09-103-840A-2
9	60.2	2.3	4411529 3	US-09-103-840A-1
10	59.6	2.3	4257 2	US-08-690-473-1
11	59.6	2.3	4257 3	US-09-259-821A-1
12	59.6	2.3	4257 3	US-08-843-659-1
13	59.6	2.3	12001 1	US-08-458-568A-11
14	58.6	2.3	1028 3	US-08-118-200-1
15	58.6	2.3	1028 3	US-08-458-745-1
16	58.2	2.3	1236 1	US-08-440-856A-1
17	58.2	2.3	471 4	US-09-252-991A-12181
18	58.2	2.3	531 4	US-09-252-991A-12413
19	58.2	2.3	567 4	US-09-252-991A-12364
20	58.2	2.3	747 4	US-09-252-991A-12243
21	58.2	2.3	2052 4	US-09-252-991A-12103
22	58.2	2.3	15378 3	US-08-785-420-1
23	57.6	2.2	4776 2	US-08-852-401-1
24	57.4	2.2	525 4	US-09-252-991A-14501
25	57.4	2.2	1092 4	US-09-252-991A-14759
26	57.4	2.2	2145 4	US-09-252-991A-14638
27	57.4	2.2	2487 4	US-09-252-991A-14985

28	57	2.2	44377 2	US-08-804-227C-7	Sequence 7, Appl
29	57	2.2	44377 2	US-08-804-198-1	Sequence 1, Appl
30	56	2.2	44377 2	US-08-804-227C-7	Sequence 7, Appl
31	56	2.2	44377 2	US-08-804-198-1	Sequence 1, Appl
32	55.8	2.2	47981 4	US-09-679-279-1	Sequence 1, Appl
33	55.2	2.1	34094 4	US-09-252-034-1	Sequence 1, Appl
34	55	2.1	2481 4	US-09-894-998A-35	Sequence 35, Appl
35	55	2.1	2721 6	5215881-2	Patent No. 5215881
36	55	2.1	4257 2	US-08-690-473-1	Sequence 1, Appl
37	55	2.1	4257 3	US-09-259-821A-1	Sequence 1, Appl
38	55	2.1	4257 3	US-08-843-659-1	Sequence 1, Appl
39	55	2.1	12001 1	US-08-458-568A-11	Sequence 11, Appl
40	54.8	2.1	3946 3	US-09-083-351-1	Sequence 1, Appl
41	54.8	2.1	3946 3	US-09-083-352-1	Sequence 1, Appl
42	54.6	2.1	936 4	US-09-416-509C-2	Sequence 2, Appl
43	54.6	2.1	2189 4	US-08-846-020A-1	Sequence 1, Appl
44	54.6	2.1	2189 4	US-09-617-871-1	Sequence 1, Appl
45	54.4	2.1	2312 1	US-07-736-176C-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-077-354B-1  
Sequence 1, Application US/09077354B  
Patent No. 6255096  
GENERAL INFORMATION:  
APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HANISH STEELE;  
APPLICANT: WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART  
TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
TITLE OF INVENTION: N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSER: SCULLY, SCOTT, MURPHY & PRESSER  
STREET: 400 GARDEN CITY PLAZA  
CITY: GARDEN CITY  
STATE: NEW YORK  
COUNTRY: UNITED STATES  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/077,354B  
FILING DATE: 22-APRIL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00747  
FILING DATE: 22-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: POKALSKY, ANN R.  
REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 12416  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516 742 4343  
TELEFAX: 516 742 4366  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2575 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORGANISM: Homo sapiens  
TISSUE TYPE: Peripheral Blood  
CELL TYPE: Leukocyte  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 102..2330  
US-09-077-354B-1

Query Match 100.0%; Score 2575; DB 3; Length 2575;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CCCGGGCTTAAAGCTTCCGAGTCCAGTGGCCGAGCCCGCAGCTGATTTGAAACGGGGCCG 60  
 Db 1 CCCGGGCTTAAAGCTTCCGAGTCCAGTGGCCGAGCCCGCAGCTGATTTGAAACGGGGCCG 60  
 61 CCCCAACCCCTGGCCGTGGCGGAGACCCGAGGACTGAGACCATGAGAGCGGTGGCGGTG 120  
 Db 61 CCCCAACCCCTGGCCGTGGCGGAGACCCGAGGACTGAGACCATGAGAGCGGTGGCGGTG 120  
 61 CCCCAACCCCTGGCCGTGGCGGAGACCCGAGGACTGAGACCATGAGAGCGGTGGCGGTG 120  
 Db 61 CCCCAACCCCTGGCCGTGGCGGAGACCCGAGGACTGAGACCATGAGAGCGGTGGCGGTG 120  
 121 CCGCGCGGCTGGGAGGTCTTCTCTGCGCGGAGCCCGGAGCGCGGAGCGGAGCGGAGCC 180  
 Db 121 CCGCGCGGCTGGGAGGTCTTCTCTGCGCGGAGCCCGGAGCGCGGAGCGGAGCGGAGCC 180  
 181 GGGAGGCGGCGCGCGCTGGCGGAGCGCTGCTGCGCGGAGCGCGGAGCGCGGAGCGGAGCC 240  
 Db 181 GGGAGGCGGCGCGCGCTGGCGGAGCGCTGCTGCGCGGAGCGCGGAGCGCGGAGCGGAGCC 240  
 241 ACTTCTCCGCTGCTGAGGAGCGCGCTGCTGCGGAGCGCGGAGCGCGGAGCGGAGCGGAGCC 300  
 Db 241 ACTTCTCCGCTGCTGAGGAGCGCGCTGCTGCGGAGCGCGGAGCGCGGAGCGGAGCGGAGCC 300  
 241 ACTTCTCCGCTGCTGAGGAGCGCGCTGCTGCGGAGCGCGGAGCGCGGAGCGGAGCGGAGCC 300  
 Db 241 ACTTCTCCGCTGCTGAGGAGCGCGCTGCTGCGGAGCGCGGAGCGCGGAGCGGAGCGGAGCC 300  
 301 TGGGCGGCGGCG 360  
 Db 301 TGGGCGGCGGCG 360  
 361 CCGGCGGCTGACCGCTACCTGCTGCGGAGCTTCTGCTGCGGAGCGGAGCGGAGCGGAGCC 420  
 Db 361 CCGGCGGCTGACCGCTACCTGCTGCGGAGCTTCTGCTGCGGAGCGGAGCGGAGCGGAGCC 420  
 421 AGCTGGCGCTGCG 480  
 Db 421 AGCTGGCGCTGCG 480  
 481 ACGAGTACCGCTATTTACGAGATGCTGAGACGAGGAGCTGAGGAGCTGAGGAGCTGAGGAG 540  
 Db 481 ACGAGTACCGCTATTTACGAGATGCTGAGACGAGGAGCTGAGGAGCTGAGGAGCTGAGGAG 540  
 541 GGGCGCGCTGGGAGCGAGAGATGAGCTGAGTGGCGCTGAGTGGAGTGAACCTGGGAGCTGG 600  
 Db 541 GGGCGCGCTGGGAGCGAGAGATGAGCTGAGTGGCGCTGAGTGGAGTGAACCTGGGAGCTGG 600  
 601 CCGGAGCGCGCGCGAGGAGGAGCTGAGGAGCGGAGTGAACCTGGGAGCTGAGGAGCTGG 660  
 Db 601 CCGGAGCGCGCGCGAGGAGGAGCTGAGGAGCGGAGTGAACCTGGGAGCTGAGGAGCTGG 660  
 661 CAGAGTCAATGAGTCTTACTGCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
 Db 661 CAGAGTCAATGAGTCTTACTGCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
 721 TGGACACTCTGGAGTGGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
 Db 721 TGGACACTCTGGAGTGGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
 781 ACCGAGGCTGAGACGAGATGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
 Db 781 ACCGAGGCTGAGACGAGATGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
 841 GGGATGTTCCCGAGGCTGTCACCAAGGAGTGGTCCCTCAAGTCAATGTCACGAGAGTGGAG 900  
 Db 841 GGGATGTTCCCGAGGCTGTCACCAAGGAGTGGTCCCTCAAGTCAATGTCACGAGAGTGGAG 900  
 901 GTTGGGGGCACTTTAACTGTTCTACTGCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTG 960  
 Db 901 GTTGGGGGCACTTTAACTGTTCTACTGCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTG 960  
 961 TATTTCCCATCATCGGAGGCTCTTCTGCTGAGAGCTGATCAAAAGATTGGACAGAGCC 1020  
 Db 961 TATTTCCCATCATCGGAGGCTCTTCTGCTGAGAGCTGATCAAAAGATTGGACAGAGCC 1020

1021 ACATCTATGAGGCGGACACTTTCAATGAGATGAGCCACCTTCTCTGAGAGCCCTCTTACC 1080  
 Db 1021 ACATCTATGAGGCGGACACTTTCAATGAGATGAGCCACCTTCTCTGAGAGCCCTCTTACC 1080  
 1081 TTGCCCCAGGACCACTGCGGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140  
 Db 1081 TTGCCCCAGGACCACTGCGGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140  
 1141 GGGCTGCTCCAGAGGCTGCTTCTTCCAGACACAGCCGAGTCTGGGGGCGCCCGCAGATGA 1200  
 Db 1141 GGGCTGCTCCAGAGGCTGCTTCTTCCAGACACAGCCGAGTCTGGGGGCGCCCGCAGATGA 1200  
 1201 GGGCTGCTCCAGAGGCTGCTTCTTCCAGACACAGCCGAGTCTGGGGGCGCCCGCAGATGA 1260  
 Db 1201 GGGCTGCTCCAGAGGCTGCTTCTTCCAGACACAGCCGAGTCTGGGGGCGCCCGCAGATGA 1260  
 1261 GCGAGCTGCTGATACCCGAGCTGCTTCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320  
 Db 1261 GCGAGCTGCTGATACCCGAGCTGCTTCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320  
 1321 TGGACAACTTTGGGGGAGAACATGAGTCTTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380  
 Db 1321 TGGACAACTTTGGGGGAGAACATGAGTCTTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380  
 1381 CAGAGGCTGCGCGCTCTTCCCACTCCACATGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAG 1440  
 Db 1381 CAGAGGCTGCGCGCTCTTCCCACTCCACATGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAG 1440  
 1441 GCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500  
 Db 1441 GCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500  
 1501 CAGTGCAGATTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560  
 Db 1501 CAGTGCAGATTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560  
 1561 ACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620  
 Db 1561 ACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620  
 1621 AGGCTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680  
 Db 1621 AGGCTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680  
 1681 CAGAGTCTGATCAACCGATCTGATGCTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740  
 Db 1681 CAGAGTCTGATCAACCGATCTGATGCTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740  
 1741 CTCCCTCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800  
 Db 1741 CTCCCTCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800  
 1801 CAGTGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860  
 Db 1801 CAGTGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860  
 1861 AGCTGGCTCTCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920  
 Db 1861 AGCTGGCTCTCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920  
 1921 ACGAG 1980  
 Db 1921 ACGAG 1980  
 1981 CAGGAG 2040  
 Db 1981 CAGGAG 2040  
 2041 CTTGTTGGGGGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100  
 Db 2041 CTTGTTGGGGGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100  
 2101 TGGCCAACTACTACACCCCTGCGTGGCGGCTTTTCTGAGAGGCGCTGTTGACAGTGTGG 2160



Db 2101 TGGCACTACTACACCTCTGCTGGCGCTTTCTTGAGAGCGCTGTGAAGATGTGG 2160  
Qy 2161 CCCAGGAGATCCCTTTCCACAGACAGATTTGCAAAATGTCTTCCACTGAGAGAG 2220  
Db 2161 CCCAGGAGATCCCTTTCCACAGACAGATTTGCAAAATGTCTTCCACTGAGAGAG 2220  
Qy 2221 CTTCTGTTCTCAGCAGCAGAGGTACCCAGCCGCGAGAGAGACACTGTGAGACTGG 2280  
Db 2221 CTTCTGTTCTCAGCAGCAGAGGTACCCAGCCGCGAGAGAGACACTGTGAGACTGG 2280  
Qy 2281 CCAAGAAAGATCTTCTCTAATATTTACCCGCTGGGTGGGCTCTGTGTATGATTC 2340  
Db 2281 CCAAGAAAGATCTTCTCTAATATTTACCCGCTGGGTGGGCTCTGTGTATGATTC 2340  
Qy 2341 GCCACACTGGGCTGTCTTCCGCTAATTCAGGAGAGATTCAGAGGCTGAGTGA 2400  
Db 2341 GCCACACTGGGCTGTCTTCCGCTAATTCAGGAGAGATTCAGAGGCTGAGTGA 2400  
Qy 2401 CAGACATCAGAGATTAACCCAGGCTGGAGAGAGCCCAAGGCTGTGTGTG 2460  
Db 2401 CAGACATCAGAGATTAACCCAGGCTGGAGAGAGCCCAAGGCTGTGTGTG 2460  
Qy 2461 ACCGCGGAGGATTTGAGAGGAAATGACCTGCTTCCACACACCAAGGTGAGATTA 2520  
Db 2461 ACCGCGGAGGATTTGAGAGGAAATGACCTGCTTCCACACACCAAGGTGAGATTA 2520  
Qy 2521 AGTACTGTTTCTTCCACTTAAAAAAGTTCAGAGGCGCGCATTC 2575  
Db 2521 AGTACTGTTTCTTCCACTTAAAAAAGTTCAGAGGCGCGCATTC 2575

## RESULT 2

US-09-077-354B-3  
Sequence 3, Application US/09077354B  
Patent No. 6255096  
GENERAL INFORMATION:  
APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;  
APPLICANT: WEBER, BIRGIT, BLANCH, LIANNE; ANSON, DONALD STEWART  
TITLE OF INVENTION: SYNTHETIC NANNALIN  
TITLE OF INVENTION: '-N-ACETYLGUCCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER  
STREET: 400 GARDEN CITY PLAZA  
CITY: GARDEN CITY  
STATE: NEW YORK  
COUNTRY: UNITED STATES  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/077.354B  
FILING DATE: 22-APRIL-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00747  
FILING DATE: 22-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: POKALSKY, ANN R.  
REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 12416  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516 742 4343  
TELEFAX: 516 742 4366  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10380 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: Chromosome 17  
FEATURE:  
NAME/KEY: exon 1  
LOCATION: 990..1372  
FEATURE:  
NAME/KEY: exon 2  
LOCATION: 2115..2262  
FEATURE:  
NAME/KEY: exon 3  
LOCATION: 3056..3202  
FEATURE:  
NAME/KEY: exon 4  
LOCATION: 3387..3472  
FEATURE:  
NAME/KEY: exon 5  
LOCATION: 5667..5923  
FEATURE:  
NAME/KEY: exon 6  
LOCATION: 7745..8955  
US-09-077-354B-3  
Query Match 55.4%; Score 1426.6; DB 3; Length 10380;  
Best Local Similarity 99.7%; Pred. No. 3.3e-289;  
Matches 1429; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1112 CATGACTCAGTGAATGATGAGGCTGTGTGCTGTGCTGCTCCAGAGCTGCTCTTCCAGACCA 1171  
Db 7734 CCGTCCACAGTGAATGATGAGGCTGTGTGCTGTGCTGCTCCAGAGCTGCTCTTCCAGACCA 7793  
Qy 1172 GCCGAGATTTGAGGAGCCGCCAGATCAGGCTGTGCTGAGAGCTGTGCTGCTGCTGCTGCTG 1231  
Db 7794 GCCGAGATTTGAGGAGCCGCCAGATCAGGCTGTGCTGAGAGCTGTGCTGCTGCTGCTGCTG 7853  
Qy 1232 CTTCTGTTCTCAGCAGCTGTTTCTGAGAGCAGGCTGTGATACCGGCACTGCTCTT 1291  
Db 7854 CTTCTGTTCTCAGCAGCTGTTTCTGAGAGCAGGCTGTGATACCGGCACTGCTCTT 7913  
Qy 1292 CCAGGCGCAGCCCTTCATCTGTGTGATGCTGTGACCACTTTGGGGGAAACATGATCTTTT 1351  
Db 7914 CCAGGCGCAGCCCTTCATCTGTGTGATGCTGTGACCACTTTGGGGGAAACATGATCTTTT 7973  
Qy 1352 TGGAGCCCTAGAGGCTGTGAAGAGAGGCGCCAGAGCTGCGGCTCTCCCACTCCAC 1411  
Db 7974 TGGAGCCCTAGAGGCTGTGAAGAGAGGCGCCAGAGCTGCGGCTCTCTCCCACTCCAC 8033  
Qy 1412 CATGTAGGCAAGGCGATGCGCCCGAGGAGCATGACCCAGAACGAAAGTGTCTATTCCT 1471  
Db 8034 CATGTAGGCAAGGCGATGCGCCCGAGGAGCATGACCCAGAACGAAAGTGTCTATTCCT 8093  
Qy 1472 CATGTAGGCTGTGCTGTGCGGCGGAGAGAGCCAGTGTGCAATTTGGAGGCTGGGATCCAG 1531  
Db 8094 CATGTAGGCTGTGCTGTGCGGCGGAGAGAGCCAGTGTGCAATTTGGAGGCTGGGATCCAG 8153  
Qy 1532 CTTTCCCGCCCGGCGATAGGAGTCTCCACCCGAGCAGAGGCGAGGCTGTGAGGCTA 1591  
Db 8154 CTTTCCCGCCCGGCGATAGGAGTCTCCACCCGAGCAGAGGCGAGGCTGTGAGGCTA 8213  
Qy 1592 GCTTCGAGTGTGTACATGCTCTCGGAGAGGCTGTGAGGCGGCAATCTGTAGCGCT 1651  
Db 8214 GCTTCGAGTGTGTACATGCTCTCGGAGAGGCTGTGAGGCGGCAATCTGTAGCGCT 8273  
Qy 1652 GGTCAAGGCGCGCTCCCTACAGATGAATACAGCATCTGTGTACACCATGTGTGT 1711  
Db 8274 GGTCAAGGCGCGCTCCCTACAGATGAATACAGCATCTGTGTACACCATGTGTGT 8333  
Qy 1712 TAGAGCTGTGCGGCTGTGCTGT 1771  
Db 8334 TAGAGCTGTGCGGCTGTGCTGT 8393

```

QY 1772 CTACAGCTGCTGAGACCTCACTCGGAGGCACTGCAAGAGCTGCTGCTACTATGA 1831
DB 8394 CTACACACCTGCTGAGACCTCACTCGGAGGCACTGCAAGAGCTGCTGCTACTATGA 8453
QY 1892 GAGAGCAGAGAGCGGCTTACTGAGCAAGAGCTGAGCTCCCTGTTGAGGCTGAGAGCGT 1891
DB 8454 GAGAGCAGAGAGCGGCTTACTGAGCAAGAGCTGAGCTCCCTGTTGAGGCTGAGAGCGT 8513
QY 1892 CCTGAGCTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1951
DB 8514 CCTGAGCTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8573
QY 1952 GCTGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2011
DB 8574 GCTGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8633
QY 2012 CTACAGAGCAAGACAGCGCTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2071
DB 8634 CTACAGAGCAAGACAGCGCTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8693
QY 2072 CTATGCTCAACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2131
DB 8694 CTATGCTCAACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8753
QY 2132 TTCTCTGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2191
DB 8754 TTCTCTGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8813
QY 2192 TGACAAAAATGCTTTCCTCACTGAGAGAGCTTCTGCTGCTGCTGCTGCTGCTGCT 2251
DB 8814 TGACAAAAATGCTTTCCTCACTGAGAGAGCTTCTGCTGCTGCTGCTGCTGCTGCT 8873
QY 2252 CCAGCGCGAGAGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2311
DB 8874 CCAGCGCGAGAGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8933
QY 2312 CTGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2371
DB 8934 CTGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8993
QY 2372 CAGGAGAGATTCAGAGGCGCCAGAGCTGAGCAAGATCAAGATTAACAGAGCTGAG 2431
DB 8994 CAGGAGAGATTCAGAGGCGCCAGAGCTGAGCAAGATCAAGATTAACAGAGCTGAG 9053
QY 2432 GAGGCGCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2491
DB 9054 GAGGCGCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9113
QY 2492 CTCACACCAACCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2544
DB 9114 CTCACACCAACCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9166

```

# RESULT 3 US-08-242-677-1

```

; Sequence 1, Application US/08242677
; Patent No. 5677143
; GENERAL INFORMATION:
; APPLICANT: Gaynor, Richard B
; APPLICANT: Wu, Foon W.
; TITLE OF INVENTION: Cellular Nucleic Acid Binding Protein
; TITLE OF INVENTION: and Uses Thereof in regulating Gene Expression and in the
; TITLE OF INVENTION: Treatment of AIDS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/242,677
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mayfield, Denise L.
; REGISTRATION NUMBER: 33,732
; REFERENCE/DOCKET NUMBER: UTSD:401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5173 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4863
; US-08-242-677-1

```

Query March 2,7%; Score 70.4; DB 1; Length 5173;

Best Local Similarity 47.0%; Pred. No. 7,4e-06;

Matches 218; Conservative 0; Mismatches 246; Indels 0; Gaps 0;

```

QY 105 GAGGCGGTGCGGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 164
DB 469 GAGGCGGTGCGGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 528
QY 165 GCAGGCGAGAGGCGCGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 224
DB 529 GAGGCGGTGCGGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 588
QY 225 CCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 284
DB 589 CCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648
QY 285 TTGACACCTACAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 344
DB 649 CTTGAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 708
QY 345 GAGGCGGTGCGGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 404
DB 709 GCGGAGAGAGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 768
QY 405 GCTGAGTCCGCTTCACTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 464
DB 769 CCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828
QY 465 ACCGAGCGCAGCGCGCAACAGATACCGCTATTACAGATGCTGACAGCAAGACTACTCC 524
DB 829 GAGCGCGTGAAGCGCGCAAGCGCGCGCTACCTGTGAGAGGCGGCGGAGGCTGTGCGCG 888
QY 525 TTCTGCTGCTGAGAGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 568
DB 889 GAGCTGCGGCGCGCGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 932

```

## RESULT 4

```

US-08-440-856A-9
; Sequence 9, Application US/08440856A
; Patent No. 5750873
; GENERAL INFORMATION:
; APPLICANT: DELAPORTA, STEPHEN L.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
; TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESSES:

```



Matches 187; Conservative 0; Mismatches 188; Indels 2; Gaps 1;

```

QY 35 GCGGCGACCTGATTGAGACGGGGCGCCGCAACCCCTTGCGCGGAGACCCGCAAGAC 94
Db 3933833 GCGGCGACCTGATTGAGACGGGGCGCCGCAACCCCTTGCGCGGAGACCCGCAAGAC 3933832
QY 95 TGAACCATGAGAGGGGTGGCGGTGGCGCGCGCGGTGGCGGTCTTCTTCTGCGCGGAGC 154
Db 3933893 GCGGCGACCTGATTGAGACGGGGCGCCGCAACCCCTTGCGCGGAGACCCGCAAGAC 3933892
QY 155 CCGGGGCGCGGAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 214
Db 3933953 TCGCGCGGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934012
QY 215 GCTGCTGGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 274
Db 3934013 GCGGATGGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934072
QY 275 CAAGCGCGG--CTTGAACATCTACGCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 332
Db 3934073 CAAGCGCGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934132
QY 333 CGCGGCTCCACGCGGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 392
Db 3934133 GCGGATGGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934192
QY 393 GCGTCCACGCTGGCGCTG 409
Db 3934193 GCGGATGGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934209

```

# RESULT 7

```

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

Query Match 2.4%; Score 61; DB 3; Length 4403765;

Best Local Similarity 49.1%; Pred. No. 0.0031;

Matches 189; Conservative 0; Mismatches 195; Indels 1; Gaps 1;

```

QY 94 CTGAGACCATGAGAGGGGTGGCGGTGGCGCGCGGTGGCGGTCTTCTTCTTCTTCTTCTT 153
Db 625458 CTTCGCGGAGCGGAGCGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCG 625517
QY 154 CCGGGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 213
Db 625518 CCGGTGGTTCGCGAAGCGGTGGCGCGCGGTGGCGCGGTGGCGCGGTGGCGCGGTGG 625577
QY 214 GCGTCTGGGCGGAGCGGCGCGGCGGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 273
Db 625578 CGGCATCGGTGGAGCGGCGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGAG 625637

```

```

QY 274 CCAAGCGGCGCTTGAACATCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 333
Db 625638 CGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625697
QY 334 GCGGCGTCCACGCGGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 393
Db 625698 TGGCGACGCGGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625756
QY 394 GCTGCCACGTGGCTTGCTGGCGGTCTTCAAGCTTGGCGGTGGCGGTGGCGGTGGCGGT 453
Db 625757 ACAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625816
QY 454 CCGGCGGAGCTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 478
Db 625817 GCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625841

```

## RESULT 8

```

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

Query Match 2.3%; Score 60.2; DB 3; Length 4403765;

Best Local Similarity 54.8%; Pred. No. 0.0045;

Matches 143; Conservative 0; Mismatches 113; Indels 5; Gaps 1;

```

QY 109 CCGTGGCGGTGGCGCGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGG 168
Db 337256 CCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 337297
QY 169 GCGACGAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 228
Db 337196 GCTCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 337137
QY 229 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 283
Db 337136 GCGGCGAGCGGAGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 337077
QY 284 CTGAGACCATGAGAGGGGTGGCGGTGGCGCGCGGTGGCGGTCTTCTTCTTCTTCTTCTT 343
Db 337076 CACTTGGCGGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 337017
QY 344 GCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 364
Db 337016 CCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336996

```

## RESULT 9

```

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:

```



QY 27 GCGCGAGGCGGCGAGCTGATTGAGACGGGGCGGCGCCACCCCTGCGCTGCGGGAGC 86  
DB 2045 GCGCGGTGCGCCCTCAGAGTGCCTGCGCGCTGCGGGATCCTGAGAGCGCTGAGCCAGG 2104  
QY 87 GCGAGACTGAGACCATGAGAGCGGTGAGCGGTGAGCGCGGTGAGGGATCTTCTCTG 146  
DB 2105 GCTTGAACGAGCACCTGCGCGCGCTGCGGGGCTGAGCGGGGCGCGCGCGAGCCG 2163  
QY 147 GCGGAGCGCGGGCGCGGCGAGCGAGCGAGCGCGGGAGCGCGCGCGCGCGCGCTC 206  
DB 2164 GCGCGCGCGAGAGAGAGCGCGCGCGCGCGCGCTTCCCGCGCGCGCGAGCGAGCGCGC 2223  
QY 207 GAGCGCGCGCTGCTGAGAGCGAGCGCGCGCGCGCGAGCTTCTCGATGCGAGAGCGCT 266  
DB 2224 GCGCGTGCAGCGGTGAGCGAGCGAGCTGCGGTGCGCGAGCGCGCGCTGCTGCTGCGC 2283  
QY 267 GCGCGTGCAGCGCGGTGAGCGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCTG 326  
DB 2284 GCGCGCGAGAGCGGTGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2343  
QY 327 GCGCGTGCAGCGCGGTGAGCGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCTG 386  
DB 2344 GCGCGTGCAGCGCGGTGAGCGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCTG 2403  
QY 387 TCTGTGCTGCGACGCTGAGCTGAGCTGCGCGCTGCGCGCTGCGCGCGCGCGCGCTG 446  
DB 2404 TCGCTG 2463  
QY 447 GCGGTGCGCG 456  
DB 2464 GCGCGCGCGG 2473

RESULT 12  
US-08-843-659-1  
Sequence 1, Application US/08843659  
Patent No. 6218103  
GENERAL INFORMATION:  
APPLICANT: Leopardi, Rosario  
APPLICANT: Roizman, Bernard  
TITLE OF INVENTION: HERPES SIMPLEX VIRUS US3 AND IC4 AS  
TITLE OF INVENTION: INHIBITORS OF APOPTOSIS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/843,659  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: ARS:519  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4257 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-843-659-1

Query Match 2.3%; Score 59.6; DB 3; Length 4257;  
Best Local Similarity 47.7%; Pred. No. 0.0013;  
Matches 205; Conservative 0; Mismatches 224; Indels 1; Gaps 1;

QY 27 GCGCGAGGCGGCGAGCTGATTGAGAGCGGCGCGCGCCACCCCTGCGCTGCGGGAGC 86  
DB 2045 GCGCGGTGCGCCCTCAGAGTGCCTGCGCGCTGCGGGATCCTGAGAGCGCTGAGCCAGG 2104  
QY 87 GCGAGACTGAGACCATGAGAGCGGTGAGCGGTGAGCGCGGTGAGGGATCTTCTCTG 146  
DB 2105 GCTTGAACGAGCACCTGCGCGCGCTGCGGGGCTGAGCGGGGCGCGCGCGAGCCG 2163  
QY 147 GCGGAGCGCGGGCGCGGCGAGCGAGCGAGCGCGGGAGCGCGCGCGCGCGCGCTC 206  
DB 2164 GCGCGCGCGAGAGAGAGCGCGCGCGCGCGCGCTTCCCGCGCGCGCGAGCGAGCGCGC 2223  
QY 207 GAGCGCGCGCTGCTGAGAGCGAGCGCGCGCGCGCGAGCTTCTCGATGCGAGAGCGCT 266  
DB 2224 GCGCGTGCAGCGGTGAGCGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCTG 326  
QY 327 GCGCGTGCAGCGCGGTGAGCGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCTG 386  
DB 2284 GCGCGTGCAGCGCGGTGAGCGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCTG 2403  
QY 387 TCTGTGCTGCGACGCTGAGCTGAGCTGCGCGCTGCGCGCTGCGCGCGCGCGCGCTG 446  
DB 2404 TCGCTG 2463  
QY 447 GCGGTGCGCG 456  
DB 2464 GCGCGCGCGG 2473

RESULT 13  
US-08-458-568A-11/C  
Sequence 11, Application US/08458568A  
Patent No. 5821339  
GENERAL INFORMATION:  
APPLICANT: Schaffer, Priscilla A.  
APPLICANT: Yeh, Lily  
TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus  
TITLE OF INVENTION: Infections  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339-15  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,568A  
FILING DATE: 02-JUNE-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/065,146  
FILING DATE: 05-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Leary Ph.D., Kathryn R.  
REGISTRATION NUMBER: 36,317  
REFERENCE/DOCKET NUMBER: DFCI-0029  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Herpes simplex virus  
STRAIN: Herpes Simplex Virus Type 1  
US-08-458-568A-11

Query March 2.3%; Score 59.6; DB 1; Length 12001;  
Best Local Similarity 47.7%; Pred. No. 0.0016;  
Matches 205; Conservative 0; Mismatches 224; Indels 1; Gaps 1;

QY 27 GCGCGAGGCGCGGAGCTGATTGACGCGGAGCCGCCACCCCTGCGCGGAGACC 86  
Db GCGCGAGGCGCGGAGCTGATTGACGCGGAGCCGCCACCCCTGCGCGGAGACC 4570  
QY 87 GCGAGAGCTGAGACCATGAGAGCGCGTGGCGCGCGCGCGTGGCGGTCTTCTG 146  
Db GCTTGAAGCGGAGACCTGAGAGCGCGTGGCGCGCGCGCGTGGCGGTCTTCTG 4511  
QY 147 GCGCGGAGCGCGGAGCGCGGAGCGAGCGAGCGCGCGCGCGCGCGCGCGCTG 206  
Db CCGCGGAGCGGAGCGCGGAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCC 4451  
QY 207 GTGCGCGCGCGCTGCTGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCT 266  
Db CGCGCTGCGCGCGCTGCTGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCTATGCGC 4391  
QY 267 CTGCGCTCCAGCGCGGCTTGAACACCTGAGCGCGCGCGCGCGCGCGCGCGCTG 326  
Db CTGCGCGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4331  
QY 327 CGGCTGCGCGCGCTGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCA 386  
Db GTGAGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 4271  
QY 387 TTCTGTGCTGCTGCA 446  
Db TCCGCTG 4211  
QY 447 GCGCGTGCAG 456  
Db GCGCGCGCGCG 4201

RESULT 14  
US-08-118-200-1  
Sequence 1, Application US/08118200  
Patent No. 6197500  
GENERAL INFORMATION:  
APPLICANT: SUTHERLAND, Grant R  
APPLICANT: RICHARDS, Robert I  
APPLICANT: SCHLESSINGER, David  
APPLICANT: NAGARAJA, Ramalath  
APPLICANT: KREMER, Eric J  
APPLICANT: YU, Sui  
APPLICANT: BAKER, Elizabeth  
APPLICANT: MULLEY, John C  
APPLICANT: MANDEL, Jean-Louis  
APPLICANT: PRITCHARD, Melanie April  
APPLICANT: LINC, Michael  
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
City: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/118,200  
FILING DATE: 09-SEP-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/802,650  
FILING DATE: 05-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,232  
FILING DATE: 20-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/638,518  
FILING DATE: 04-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/966,517  
FILING DATE: 23-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 020160-164  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1028 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-118-200-1  
Query March 2.3%; Score 58.6; DB 3; Length 1028;  
Best Local Similarity 59.2%; Pred. No. 0.0015;  
Matches 100; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 70 CTGCGCTGTCGCGGAGCCGCGAGACTGAGACCATGAGAGCGGCTGCGCGCGCG 129  
Db CCGGCGGAGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 399  
QY 130 TGGGGGCTCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 189  
Db CCGCGCGGCTGCG 459  
QY 190 CGGCGCTGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 238  
Db CCGCGCGGC 508

RESULT 15  
US-08-458-745-1  
Sequence 1, Application US/08458745  
Patent No. 6242576  
GENERAL INFORMATION:  
APPLICANT: SUTHERLAND, Grant R  
APPLICANT: RICHARDS, Robert I  
APPLICANT: SCHLESSINGER, David  
APPLICANT: NAGARAJA, Ramalath  
APPLICANT: KREMER, Eric J  
APPLICANT: YU, Sui  
APPLICANT: BAKER, Elizabeth  
APPLICANT: MULLEY, John C



APPLICANT: MANDEL, Jean-Louis  
APPLICANT: PITCHARD, Melanie April  
APPLICANT: LYNCH, Michael  
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED  
TITLE OF INVENTION: FRAGILE X SYNDROME  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,745  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/118,200  
FILING DATE: 09-SEP-1993  
APPLICATION NUMBER: US 07/802,650  
FILING DATE: 05-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,232  
FILING DATE: 20-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/638,518  
FILING DATE: 04-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/966,517  
FILING DATE: 23-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feuyl, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 020160-164  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1028 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-458-745-1

Query Match 2.3%; Score 58.6; DB 3; Length 1028;  
Best Local Similarity 59.2%; Pred. No. 0.0015;  
Matches 100; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
QY 70 CTGGCGGTGGCGGAGCCCGGAGCATGAGAGCGGTGGCGCGCGCGG 129  
DB 340 CGGCGCGAGCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 399  
QY 130 TGGGAGTCTTCTCTGCGCGGCGCGCGCGCGCGCGCGCGCGAGCGAGCGCCGCGAGGCGG 189  
DB 400 CGGCGCGCGGTGGCG 459  
QY 190 CGGCGCGTGGCG 238  
DB 460 CGGCG 508

Search completed: February 13, 2004, 13:34:51  
Job time : 179 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2004, 12:03:44 ; Search time 890 Seconds

(without alignments)  
10657.701 Million cell updates/sec

Title: US-09-836-613-1

Perfect score: 2575  
Sequence: 1 CCCGGGCTTACCTTCGGGT.....AGTCAGCCGCCGCAATTC 2575

Scoring table: IDENTITY NUC  
Gap 10.0, Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCN\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCNUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2575	100.0	2575	11	US-09-836-613-1
2	1426.6	55.4	10380	11	US-09-836-613-3
3	1211	47.0	1211	13	US-10-029-386-22784
4	485.6	18.9	2796	13	US-10-027-632-111960
5	485.6	18.9	2796	13	US-10-027-632-111960
6	456.6	17.7	808	14	US-10-027-632-141597
7	456.6	17.7	808	14	US-10-027-632-141597
8	445.4	17.3	499	11	US-09-918-995-30489
9	426	16.5	522	13	US-10-029-386-9071
10	395	15.3	449	11	US-09-918-995-24382
11	358.4	13.9	468	11	US-09-918-995-21830
12	183.4	7.1	2184	15	US-10-156-761-2002
13	183.4	7.1	2184	15	US-10-156-761-1
14	155	6.0	3114	15	US-10-156-761-5965
15	155	6.0	9025608	15	US-10-156-761-1

16	88	3.4	2202	12	US-10-260-238-900	Sequence 900, App
17	73.6	2.9	11817	15	US-10-156-761-2884	Sequence 2884, App
18	73.6	2.9	125746	15	US-10-156-761-15102	Sequence 15102, A
19	72.6	2.8	1011	13	US-10-029-386-25041	Sequence 25041, A
20	71	2.8	4020	11	US-09-796-679-5	Sequence 5, Appl1
21	70.4	2.7	5173	15	US-10-880-107-1356	Sequence 3356, App
22	70.4	2.7	5173	15	US-10-171-561-159	Sequence 159, App
23	70	2.7	10144	12	US-09-880-107-2168	Sequence 2168, App
24	69.2	2.7	51657	12	US-10-057-475B-10475	Sequence 10475, A
25	69.2	2.7	51657	12	US-10-154-884B-10475	Sequence 48, Appl1
26	68.6	2.7	2561	10	US-09-976-740-48	Sequence 48, Appl1
27	68.6	2.7	2561	12	US-10-616-187-48	Sequence 48, Appl1
28	68.6	2.7	2561	14	US-10-023-523-48	Sequence 48, Appl1
29	68.6	2.7	2561	14	US-10-023-523-48	Sequence 48, Appl1
30	68.2	2.6	2271	13	US-10-101-510-133	Sequence 133, App
31	68.2	2.6	2487	15	US-10-205-823-135	Sequence 15, App
32	66.2	2.6	152331	14	US-10-095-407-16	Sequence 16, Appl1
33	64.4	2.5	1614	10	US-09-976-740-45	Sequence 45, Appl1
34	64.4	2.5	1614	12	US-10-616-187-45	Sequence 45, Appl1
35	64.4	2.5	1614	14	US-10-023-523-45	Sequence 45, Appl1
36	64.4	2.5	1614	14	US-10-023-523-45	Sequence 45, Appl1
37	64.4	2.5	3957	13	US-10-200-562-193	Sequence 193, App
38	64.4	2.5	3957	13	US-10-237-551-193	Sequence 13, Appl1
39	64.4	2.5	4225	13	US-10-141-541-3	Sequence 50, Appl1
40	64.4	2.5	12425	10	US-09-976-740-50	Sequence 50, Appl1
41	64.4	2.5	12425	12	US-10-616-187-50	Sequence 50, Appl1
42	64.4	2.5	12425	14	US-10-023-523-50	Sequence 50, Appl1
43	64.4	2.5	12425	14	US-10-023-523-50	Sequence 50, Appl1
44	64.4	2.5	154746	13	US-09-827-668-8	Sequence 8, Appl1
45	64.4	2.5	154746	13	US-09-827-668-8	Sequence 8, Appl1

## ALIGNMENTS

RESULT 1  
US-09-836-613-1  
Sequence 1, Application US/09836613  
Publication No. US20030039643A1

GENERAL INFORMATION:

APPLICANT: HOBWOOD, JOHN JOSEPH; SCOTT, HAMISH STEWART;  
WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART

TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES

ENCODING SAME

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:  
ADDRESSER: NIXON PEABODY LLP

STREET: 990 STEWART AVENUE  
CITY: GARDEN CITY

STATE: NEW YORK  
COUNTRY: UNITED STATES

ZIP: 11530

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/836,613

FILING DATE: 17-Apr-2001

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00747

FILING DATE: 22-NOV-1996

ATTORNEY/AGENT INFORMATION:  
NAME: POKALSKY, ANN R.

REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 2249/104

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516 742 4366

TELEFAX: 516 742 4366

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 2575 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Peripheral Blood  
CELL TYPE: Leukocyte  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 102..2330  
SEQUENCE DESCRIPTION: SEQ ID NO: 1  
US-09-836-613-1

Query Match 100.0%; Score 2575; DB 11; Length 2575;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCGGGCTTACCTTCGGGTCACGTGGCCGAGGCCGAGCTGATTGAGCGCGGCGG 60  
DB 1 CCCGGGCTTACCTTCGGGTCACGTGGCCGAGGCCGAGCTGATTGAGCGCGGCGG 60  
QY 61 CCCCAACCCCTGGCCGCTGCGGAGACCCGAGACTGAGACATGAGAGCGGTCGTCG 120  
DB 61 CCCCAACCCCTGGCCGCTGCGGAGACCCGAGACTGAGACATGAGAGCGGTCGTCG 120  
QY 121 CCGCGGCGGTGGGGGTCCTTCTCTGGCCGAGGCGCGGAGCGGAGCGGAGCGGCG 180  
DB 121 CCGCGGCGGTGGGGGTCCTTCTCTGGCCGAGGCGCGGAGCGGAGCGGAGCGGCG 180  
QY 181 GGGAGGCGGCG 240  
DB 181 GGGAGGCGGCG 240  
QY 241 ACTTCTCGGTGCGGTGAGAGCGCGCTGCTGCTGCAAGCGGAGCTTGAACACTTACAG 300  
DB 241 ACTTCTCGGTGCGGTGAGAGCGCGCTGCTGCTGCAAGCGGAGCTTGAACACTTACAG 300  
QY 301 TGGGCG 360  
DB 301 TGGGCG 360  
QY 361 CGGGGCTGCAACCGCTACCTGCGGACTTCTGTGGCTGCCAGCTGGCTGGTCCGCGCT 420  
DB 361 CGGGGCTGCAACCGCTACCTGCGGACTTCTGTGGCTGCCAGCTGGCTGGTCCGCGCT 420  
QY 421 AGCTGCGCTGCG 480  
DB 421 AGCTGCGCTGCG 480  
QY 481 ACAGGTACCGGTATTACAGATGTGTGACAGCAAGACTCTCTTCTGTGTGTGGAGCT 540  
DB 481 ACAGGTACCGGTATTACAGATGTGTGACAGCAAGACTCTCTTCTGTGTGTGGAGCT 540  
QY 541 GGGCGCGCTGGAGCGAGATAGATGATGCGCTGATGAGCTCAACTGGCACTGG 600  
DB 541 GGGCGCGCTGGAGCGAGATAGATGATGCGCTGATGAGCTCAACTGGCACTGG 600  
QY 601 CTTGGAGCGCGCGAGAGGCGATCTGGAGCGGAGTGAACCTGGGCTTGGGCTGACCA 660  
DB 601 CTTGGAGCGCGCGAGAGGCGATCTGGAGCGGAGTGAACCTGGGCTTGGGCTGACCA 660  
QY 661 CAGAGATCAATGATTTCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
DB 661 CAGAGATCAATGATTTCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
QY 721 TGGACACCTGGGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
DB 721 TGGACACCTGGGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
QY 781 ACCGGGCTCTGACACAGATGCGCTCTTGGGCAATGACCCAGTGTGCTGCTGCTGCG 840

DB 781 ACCGGGCTCTGACACAGATGCGCTCTTGGGCAATGACCCAGTGTGCTGCTGCG 840  
QY 841 GGCATGTTCCCGAGGCTGTACACAGGCTGTTCCTCAGCTCAATGTACAGAGATGG 900  
DB 841 GGCATGTTCCCGAGGCTGTACACAGGCTGTTCCTCAGCTCAATGTACAGAGATGG 900  
QY 901 GTTGGGCGCACTTTAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 960  
DB 901 GTTGGGCGCACTTTAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 960  
QY 961 TATTCCCATATGAGAGGCTCTTCTGAGAGGCTGATTAAGAGTTGGCAGAGCC 1020  
DB 961 TATTCCCATATGAGAGGCTCTTCTGAGAGGCTGATTAAGAGTTGGCAGAGCC 1020  
QY 1021 ACATCTATGGGCGGACACTTTCAATGATGAGAGCCAGCTTCTCTAGAGCCCTTACC 1080  
DB 1021 ACATCTATGGGCGGACACTTTCAATGATGAGAGCCAGCTTCTCTAGAGCCCTTACC 1080  
QY 1081 TTGCGGAGCCACACATGCGCTTATAGGCGATGAGTGAATGAGAGGCTGTGT 1140  
DB 1081 TTGCGGAGCCACACATGCGCTTATAGGCGATGAGTGAATGAGAGGCTGTGT 1140  
QY 1141 GACTGCTCAAGGCTGCTCTTCCAGACACAGCCGAGTTCTGGGGCCGCGCCAGATCA 1200  
DB 1141 GACTGCTCAAGGCTGCTCTTCCAGACACAGCCGAGTTCTGGGGCCGCGCCAGATCA 1200  
QY 1201 GGGCTGTGCTGGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260  
DB 1201 GGGCTGTGCTGGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260  
QY 1261 GCGAGCTGTATACCGCGCACTGCTCTTCCAGAGGCGAGCCCTTCACTGTGTCATG 1320  
DB 1261 GCGAGCTGTGTATACCGCGCACTGCTCTTCCAGAGGCGAGCCCTTCACTGTGTCATG 1320  
QY 1321 TGCACACTTTGGGGGAAACCATGCTTTTGGAGCCCTTAGAGCTGTGAGCGAGGCC 1380  
DB 1321 TGCACACTTTGGGGGAAACCATGCTTTTGGAGCCCTTAGAGCTGTGAGCGAGGCC 1380  
QY 1381 CAGAGCTGCGCGCTCTTCCCAATCCACATGATAGGAGAGGCGAGCCCGCGAG 1440  
DB 1381 CAGAGCTGCGCGCTCTTCCCAATCCACATGATAGGAGAGGCGAGCCCGCGAG 1440  
QY 1441 GCATGAGCAGAGAGATGATGATTTCCCTATGAGCTGAGCTGGCGTGGGAAAGACC 1500  
DB 1441 GCATGAGCAGAGAGATGATGATTTCCCTATGAGCTGAGCTGGCGTGGGAAAGACC 1500  
QY 1501 CAGTCCCAATTTGGCAAGCTTGGGTGACAGCTTTGCGCGCGCGCGGTATGGGCTCCC 1560  
DB 1501 CAGTCCCAATTTGGCAAGCTTGGGTGACAGCTTTGCGCGCGCGCGGTATGGGCTCCC 1560  
QY 1561 ACCCGGAGCGAGGCGAGCTGAGAGCTACGTCTCGGAGTGTGTACACTGCTCCG 1620  
DB 1561 ACCCGGAGCGAGGCGAGCTGAGAGCTACGTCTCGGAGTGTGTACACTGCTCCG 1620  
QY 1621 AGGCTTGAAGGGGCGCAATGTATGCGCGCTGTGAGCGGCGCTCTTACAGATGATA 1680  
DB 1621 AGGCTTGAAGGGGCGCAATGTATGCGCGCTGTGAGCGGCGCTCTTACAGATGATA 1680  
QY 1681 CGAGCATGTGTATACCACTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1740  
DB 1681 CGAGCATGTGTATACCACTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1740  
QY 1741 CTCCTCTCTGGGCAACAGCCCGCGCTTCCGCTAGACCTGCTGCTGCTGCTGCTG 1800  
DB 1741 CTCCTCTCTGGGCAACAGCCCGCGCTTCCGCTAGACCTGCTGCTGCTGCTGCTG 1800  
QY 1801 CAGTGAAGAGAGCTGATGCTTACTATGAGAGGCAAGAGGCTTACTGTAGCAAG 1860  
DB 1801 CAGTGAAGAGAGCTGATGCTTACTATGAGAGGCAAGAGGCTTACTGTAGCAAG 1860  
QY 1861 AGTGGGCTCTCTTGAAGGCTGAGAGGCTCTGCGCTATGAGCTGCTGCGGCACTGG 1920  
DB 1861 AGTGGGCTCTCTTGAAGGCTGAGAGGCTCTGCGCTATGAGCTGCTGCGGCACTGG 1920

QY 1921 ACAGAGTGTGCTAGTGAACAGCGCTCTTCTGCTGAGCTGAGTACAGCGCCGAG 1980  
DB 1921 ACAGAGTGTGCTAGTGAACAGCGCTCTTCTGCTGAGCTGAGTACAGCGCCGAG 1980  
QY 1981 CAGGCGCAGTGAAGGAGCGGCGGCTTTCTAGCAGCAAGAGCGGCTTACCACTGA 2040  
DB 1981 CAGGCGCAGTGAAGGAGCGGCGGCTTTCTAGCAGCAAGAGCGGCTTACCACTGA 2040  
QY 2041 CTTGTGGGGGCGCAAGAGCGCAATCTGAGCTTGAACAGAGCTGAGCGGCTTGA 2100  
DB 2041 CTTGTGGGGGCGCAAGAGCGCAATCTGAGCTTGAACAGAGCTGAGCGGCTTGA 2100  
QY 2101 TGCGCAACTACTACACCCCTGCTGCGGCTTTCTGAGGCGCTGTTGAAGTGTG 2160  
DB 2101 TGCGCAACTACTACACCCCTGCTGCGGCTTTCTGAGGCGCTGTTGAAGTGTG 2160  
QY 2161 CCGAGGCGATCCCTTCCAGCAGCAGTGTGAACAAATGCTTCCAACTGAGAGAG 2220  
DB 2161 CCGAGGCGATCCCTTCCAGCAGCAGTGTGAACAAATGCTTCCAACTGAGAGAG 2220  
QY 2221 CTTTCTGCTAGCAGCAGAGTACCCAGCGCGAGCGAGAGCACTGTGAGCTTGA 2280  
DB 2221 CTTTCTGCTAGCAGCAGAGTACCCAGCGCGAGCGAGAGCACTGTGAGCTTGA 2280  
QY 2281 CCAAGAAAGTCTCTCTCAAAATTTTACCCCGCTGAGGCTGCTGTTGATGATTC 2340  
DB 2281 CCAAGAAAGTCTCTCTCAAAATTTTACCCCGCTGAGGCTGCTGTTGATGATTC 2340  
QY 2341 GCCACCACTGAGGCTTGTTCCTGCTAATTCAGAGGAGATTCAGAGGCGCCAGAGCTGA 2400  
DB 2341 GCCACCACTGAGGCTTGTTCCTGCTAATTCAGAGGAGATTCAGAGGCGCCAGAGCTGA 2400  
QY 2401 CAGACATCAAGATTAACCCAGGCTGAGAGAGCGCCAGCGCTGCTGAGGCTG 2460  
DB 2401 CAGACATCAAGATTAACCCAGGCTGAGAGAGCGCCAGCGCTGCTGAGGCTG 2460  
QY 2461 ACCTGGGGGAGTGGAGGAAATGACCTGCTCAGCAGCAGCAAGTGTGGATTA 2520  
DB 2461 ACCTGGGGGAGTGGAGGAAATGACCTGCTCAGCAGCAGCAAGTGTGGATTA 2520  
QY 2521 AGTACTGTTCTTCCACTTAATAAAAAAAAAAAGTCAGAGCGCGCAATTC 2575  
DB 2521 AGTACTGTTCTTCCACTTAATAAAAAAAAAAAGTCAGAGCGCGCAATTC 2575

RESULT 2  
US-09-836-613-3  
Sequence 3, Application US/09836613  
Publication No. US2003003643A1  
GENERAL INFORMATION:  
APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;  
WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART  
TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES  
ENCODING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON PEARODY LLP  
STREET: 990 STEWART AVENUE  
CITY: GARDEN CITY  
STATE: NEW YORK  
COUNTRY: UNITED STATES  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/836,613  
FILING DATE: 17-Apr-2001  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/00747  
FILING DATE: 22-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: POKALSKY, ANN R.  
REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 2249/104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516 742 4343  
TELEFAX: 516 742 4366  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10380 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: Chromosome 17  
FEATURE:  
NAME/KEY: exon 1  
LOCATION: 990..1372  
FEATURE:  
NAME/KEY: exon 2  
LOCATION: 2115..2262  
FEATURE:  
NAME/KEY: exon 3  
LOCATION: 3056..3202  
FEATURE:  
NAME/KEY: exon 4  
LOCATION: 3387..3472  
FEATURE:  
NAME/KEY: exon 5  
LOCATION: 5667..5923  
NAME/KEY: exon 6  
LOCATION: 7745..8955  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-836-613-3  
Query Match 55.4%; Score 1426.6; DB 11; Length 10380;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1429; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1112 CATGACTGAGTGAATGAGAGGCTGTGTGCTGCTCAAGGCTGCTCTTCCAGACCA 1171  
DB 7734 CTTGTCCACAGTGAATGAGAGGCTGTGTGCTGCTCAAGGCTGCTCTTCCAGACCA 7793  
QY 1172 GCGGAGTCTGAGGAGCGCGCCAGATCAGGCTGTGTGAGAGCTGTGCGCGCG 1231  
DB 7794 GCGGAGTCTGAGGAGCGCGCCAGATCAGGCTGTGTGAGAGCTGTGCGCGCG 7853  
QY 1232 CCTCTGTGTGAGACGTGTTGTCTGAGAGCAGGCTGTGTATACCCGACGCTCCT 1291  
DB 7854 CCTCTGTGTGAGACGTGTTGTCTGAGAGCAGGCTGTGTATACCCGACGCTCCT 7913  
QY 1292 CCAGGCGCAGCCCTCATCTGATGATCTGACAACTTTGGGGGAAACATGATCTTT 1351  
DB 7914 CCAGGCGCAGCCCTCATCTGATGATCTGACAACTTTGGGGGAAACATGATCTTT 7973  
QY 1352 TGAAGCCTTAAGAGCTGTGAAGAGAGGCGCCAGAACTGCGGCTCTTCCCACTCCAG 1411  
DB 7974 TGAAGCCTTAAGAGCTGTGAAGAGAGGCGCCAGAACTGCGGCTCTTCCCACTCCAG 8033  
QY 1412 CATGTGAGCAGGCGCATGAGCGCGGCGAGGCGCATGAGCAAGAGTGTATCTCCT 1471  
DB 8034 CATGTGAGCAGGCGCATGAGCGCGGCGAGGCGCATGAGCAAGAGTGTATCTCCT 8093  
QY 1472 CATGCTGAGCTGTGCGCAGAAAGACCCAGTGTCAATTTGGAGCTGTGATCAAG 1531  
DB 8094 CATGCTGAGCTGTGCGCAGAAAGACCCAGTGTCAATTTGGAGCTGTGATCAAG 8153

1532 CTTTCCGCGCGCGGCTATGAGGCTCTCCACCAGCCGAGCGAGCGGCGAGCGTGAAGCTACT 1591  
8154 CTTTCCGCGCGCGGCTATGAGGCTCTCCACCAGCCGAGCGAGCGGCGAGCGTGAAGCTACT 8213  
1592 GCTCCGAGAGTGTATCACTGCTCCGAGGAGGCTCCAGAGGCGCAATCTGATGCGCT 1651  
8214 GCTCCGAGAGTGTATCACTGCTCCGAGGAGGCTCCAGAGGCGCAATCTGATGCGCT 8273  
1652 GGTCAAGGCGCGCTCTCCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1711  
8274 GGTCAAGGCGCGCTCTCCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 8333  
1712 TGAAGCCTGAGCGCTGCTGCTCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1771  
8334 TGAAGCCTGAGCGCTGCTGCTCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8393  
1772 CTACAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1831  
8394 CTACAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8453  
1832 GAGAGCAAGAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1891  
8454 GAGAGCAAGAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8513  
1892 CTTGCGCTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1951  
8514 CTTGCGCTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8573  
1952 GCTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2011  
8574 GCTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8633  
2012 CTACAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2071  
8634 CTACAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8693  
2072 CTACAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2131  
8694 CTACAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8753  
2132 TTTCTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2191  
8754 TTTCTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8813  
2192 TGAACAAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2251  
8814 TGAACAAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8873  
2252 CCAAGCCGAGAGGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2311  
8874 CCAAGCCGAGAGGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8933  
2312 CTTGAGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2371  
8934 CTTGAGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8993  
2372 CAGGCGAGATTCAGAGGCGCGAGAGCTGAGCAGACATCAAGATTAACAGAGGCTGAGAG 2431  
8994 CAGGCGAGATTCAGAGGCGCGAGAGCTGAGCAGACATCAAGATTAACAGAGGCTGAGAG 9053  
2432 GAGGCGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2491  
9054 GAGGCGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9113  
2492 CTCAC 2544  
9114 CTCAC 9166

RESULT 3  
US-10-029-386-22784  
; Sequence 22784, Application US/10029386  
; Publication No. US20030194704A1

GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Hank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: ABOICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; NUMBER OF SEQ ID NOS: 3428  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 22784  
; LENGTH: 1211  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO U34879.1  
; OTHER INFORMATION: EXPRESSED IN HEMA, SIGNAL = 0.74  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.93  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6  
; OTHER INFORMATION: NT HIT: 514785987, EVALUATE 0.00e+00  
; OTHER INFORMATION: SWISSPROT HIT: P54802, EVALUATE 0.00e+00  
; OTHER INFORMATION: EST\_HUMAN HIT: BE410798.1, EVALUATE 0.00e+00  
US-10-029-386-22784  
Query Match 47.0%; Score 1211; DB 13; Length 1211;  
Best Local Similarity 100.0%; Pred. No. 1.5e-299; Indels 0; Gaps 0;  
Matches 1211; Conservative 0; Mismatches 0  
1123 TGAATACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1182  
1 TGAATACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60  
1183 GAGGAGCGCGCCAGATAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1242  
61 GAGGAGCGCGCCAGATAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
1243 TGAACCTGTTGCTGAGAGGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1302  
121 TGAACCTGTTGCTGAGAGGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
1303 CTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1362  
181 CTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
1363 AGGCTGAGAGGAGGCGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1422  
241 AGGCTGAGAGGAGGCGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
1423 CAGGAGATGAGCGCGCGAGAGGAGATGAGAGGAGATGAGAGGAGATGAGAGGAGATGAGAGG 1482  
301 CAGGAGATGAGCGCGCGAGAGGAGATGAGAGGAGATGAGAGGAGATGAGAGGAGATGAGAGG 360  
1483 TGGGCTGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 1542  
361 TGGGCTGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 420  
1543 GAGGCTGATGAGGCTCTCCACCGGAGCGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 1602  
421 GAGGCTGATGAGGCTCTCCACCGGAGCGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 480  
1603 TGTACAACTGCTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1662  
481 TGTACAACTGCTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
1663 GGTCCCTACAGATGATATACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1722  
541 GGTCCCTACAGATGATATACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
1723 GGTCCCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1782  
601 GGTCCCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660

```

QY 1783 TGAACCTCACTGGAGAGGAGTGCAGAGCTGTCAGCTTGTATGATGAGAGGAGAGAA 1842
DB 661 TGAACCTCACTGGAGAGGAGTGCAGAGCTGTCAGCTTGTATGATGAGAGGAGAGAA 720
QY 1843 GGGCTTACTGAGCAAGAGAGCTGGCTCCCTGTTGAGGGCTGGAGGGCTTCTGGCTATG 1902
DB 721 GGGCTTACTGAGCAAGAGAGCTGGCTCCCTGTTGAGGGCTGGAGGGCTTCTGGCTATG 780
QY 1903 AGCTGCTGCGGAGCACTGAGCAGAGTGTGAGTGTGAGCAGCGCTTCTTGGCTGGAGAGCT 1962
DB 781 AGCTGCTGCGGAGCACTGAGCAGAGTGTGAGTGTGAGCAGCGCTTCTTGGCTGGAGAGCT 840
QY 1963 GGTAGAGCAGGCGCGAGAGCAGGAGTGTGAGTGTGAGCAGCGCTTCTTGGCTGGAGAG 2022
DB 841 GGTAGAGCAGGCGCGAGAGCAGGAGTGTGAGTGTGAGCAGCGCTTCTTGGCTGGAGAG 900
QY 2023 ACAGCGGCTACCGAGCTGAGCTTGTGAGGGGCGAGAGGAGAGATCTCTGAGCTATGCGCA 2082
DB 901 ACAGCGGCTACCGAGCTGAGCTTGTGAGGGGCGAGAGGAGAGATCTCTGAGCTATGCGCA 960
QY 2083 AGCAGCTGCGGGGTTGTGAGCACTACACCCCTGCTGCGGCTTCTTGGCTGGAGG 2142
DB 961 AGCAGCTGCGGGGTTGTGAGCACTACACCCCTGCTGCGGCTTCTTGGCTGGAGG 1020
QY 2143 CGCTGCTTGAAGTGTGAGGCGCGAGGCTTCTTGGCTGGAGGAGGAGGAGGAGGAGG 2202
DB 1021 CGCTGCTTGAAGTGTGAGGCGCGAGGCTTCTTGGCTGGAGGAGGAGGAGGAGGAGG 1080
QY 2203 TCTTCAACTGAGAGGAGGCTTCTTGGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2262
DB 1081 TCTTCAACTGAGAGGAGGCTTCTTGGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
QY 2263 GAGACACTGTGAGCTTGGCAAGAGATCTTCTCAATATTACCCGCGCTGGGCTGGCGG 2322
DB 1141 GAGACACTGTGAGCTTGGCAAGAGATCTTCTCAATATTACCCGCGCTGGGCTGGCGG 1200
QY 2323 GCTCTTGGTGA 2333
DB 1201 GCTCTTGGTGA 1211

```

```

RESULT 4
US-10-027-632-111960
; Sequence 111960, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111960
; LENGTH: 2796
; TYPE: DNA
; ORGANISM: Human

```

```

US-10-027-632-111960
Query Match 18.9%; Score 485.6; DB 13; Length 2796;
Best Local Similarity 85.9%; Pred. No. 4,8e-114;
Matches 550; Conservative 0; Mismatches 89; Indels 1; Gaps 1;
QY 1 CCGGAGCTTAAAGCTTGGGTCACGTGAGCGAGAGGCGGAGAGTGAATTGAGCGGAGCGG 60
DB 1017 CCGGAGCTTAAAGCTTGGGTCACGTGAGCGAGAGGCGGAGAGTGAATTGAGCGGAGCGG 1075
QY 61 CCGGAGCTTAAAGCTTGGGTCACGTGAGCGAGAGGCGGAGAGTGAATTGAGCGGAGCGG 120
DB 1076 CCGGAGCTTAAAGCTTGGGTCACGTGAGCGAGAGGCGGAGAGTGAATTGAGCGGAGCGG 1135
QY 121 CCGGAGCTTAAAGCTTGGGTCACGTGAGCGAGAGGCGGAGAGTGAATTGAGCGGAGCGG 180
DB 1136 CCGGAGCTTAAAGCTTGGGTCACGTGAGCGAGAGGCGGAGAGTGAATTGAGCGGAGCGG 1195
QY 181 GGGAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 240
DB 1196 GGGAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1255
QY 241 ACTTCTCCGTCGAGTGAAGCGGCTCTGAGCTGCAAGCGGAGCTTGAACCTTACAGCC 300
DB 1256 ACTTCTCCGTCGAGTGAAGCGGCTCTGAGCTGCAAGCGGAGCTTGAACCTTACAGCC 1315
QY 301 TGGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 360
DB 1316 TGGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1375
QY 361 CGGAGCTGAGCAGCTACCTGCGGAGCTTGTGAGGCTGCGAGCGGAGCTGCGGAGCTGCG 420
DB 1376 CGGAGCTGAGCAGCTACCTGCGGAGCTTGTGAGGCTGCGAGCGGAGCTGCGGAGCTGCG 1435
QY 421 AGCTGCGCTGCGGAGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGG 480
DB 1436 AGCTGCGCTGCGGAGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGG 1495
QY 481 ACAGGTACCGCTATTACGAATGTGTGACAGCGCAAGCTACCTTCTGTGTGGAGCT 540
DB 1496 ACAGGTACCGCGCCCGAGCTTCCCGCGCTGCGGAGCGGAGCGGAGCTTCCCGCGAGCT 1555
QY 541 GGGCGGCTGAGAGCAGAGATAGACTGATGAGCGCTGATGAGCATCACTTGGCACTGG 600
DB 1556 GGGCGGCTGAGAGCAGAGATAGACTGATGAGCGCGTGGAGCGCTTGGAGCGGAGCGGAGCT 1615
QY 601 CCTGAGCGGCGGAGAGCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCGGAGCTGAGCT 640
DB 1616 CCTGAGCGGCGGAGAGCTGTGTGAGCGCTTGAAGCGAGCACTTGGCT 1655

```

```

RESULT 5
US-10-027-632-111960
; Sequence 111960, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28

```

```
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11960
; LENGTH: 2796
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-11960

Query Match      18.3%; Score 485.6; DB 14; Length 2796;
Best Local Similarity 85.9%; Pred. No. 4.8e-114;
Matches 550; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

QY 1 CCGGGGCTTACCTGCGGAGTCAAGTGGCCGAGGCGGAGGCTGATTGGAGCGGGGCGG 60
DB 1017 CCGGGGCTTACCTGCGGAGTCAAGTGGCCGAGGCGGAGGCTGATTGGAGCGGGGCGG 1075
QY 61 CCCCACCCCTGGCCGCTGGCGGAGACCCGAGACTGAGACATGAGAGCGGTGGCGGTGG 120
DB 1076 CCCCACCCCTGGCCGCTGGCGGAGACCCGAGACTGAGACATGAGAGCGGTGGCGGTGG 1135
QY 121 CCGCGGCGGTGGGAGTCTTCTCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 1136 CCGCGGCGGTGGGAGTCTTCTCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1195
QY 181 GGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 1196 GGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1255
QY 241 ACTTCTCCGTTGGTGGGAGAGCGGCTTGGCTGACCAAGCGGCGTGGACACTTACAGCC 300
DB 1256 ACTTCTCCGTTGGTGGGAGAGCGGCTTGGCTGACCAAGCGGCGTGGACACTTACAGCC 1315
QY 301 TGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
DB 1316 TGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1375
QY 361 CCGGCGCTGACCGCTACCTGCGGAGTCTTGTGGCTGCGACGTTGGCTTCCGCTTTC 420
DB 1376 CCGGCGCTGACCGCTACCTGCGGAGTCTTGTGGCTGCGACGTTGGCTTCCGCTTTC 1435
QY 421 AGCTGGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
DB 1436 AGCTGGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1495
QY 481 ACAAGTACCGCTATTACAGAAATGTGTGACCGCAAAAGCTACTCTTGTGTGGTGA 540
DB 1496 ACAAGTACCGCTATTACAGAAATGTGTGACCGCAAAAGCTACTCTTGTGTGGTGA 1555
QY 541 GGGCGCGCTGGGAGCGAGAGATAGACTGAGATGGCGCTGATCAACCTGGACCTGG 600
DB 1556 GCTGGCACCCCAATTCGGGAGGCTGAGGCGGAGAGCGCTGGCGGAAAGCCCAAGTGGCGG 1615
QY 601 CCGGAGCGGCGCAGAGAGGCGCATCTGGACGCGGCTGATCTT 640
DB 1616 CCGGAGCGGCGCAGAGAGGCGCATCTGGACGCGGCTGATCTT 1655

RESULT 6
US-10-027-632-141597
; Sequence 141597, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR APPLICATION NUMBER: 2002-04-30
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
```

```
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 141597
; LENGTH: 808
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-141597

Query Match      17.7%; Score 456.6; DB 13; Length 808;
Best Local Similarity 99.8%; Pred. No. 1e-106;
Matches 456; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2088 CTGGCGGGGTTGGTGGCCACTACTACACCCCTGCGTGGCGGCTTCTCTGGAGGCGCTG 2147
DB 1 CTGGCGGGGTTGGTGGCCACTACTACACCCCTGCGTGGCGGCTTCTCTGGAGGCGCTG 60
QY 2148 GTTACAGTGTGGCCAGGCGCATCCCTTCCACAGCAGCAGTGGACAAATGTCTTC 2207
DB 61 GTTACAGTGTGGCCAGGCGCATCCCTTCCACAGCAGCAGTGGACAAATGTCTTC 120
QY 2208 CACTGAGAGAGGCTTCTGTTCTAGCAGAGAGGTAACCCAGCCAGCCGAGGAGAC 2267
DB 121 CACTGAGAGAGGCTTCTGTTCTAGCAGAGAGGTAACCCAGCCAGCCGAGGAGAC 180
QY 2268 ACTGAGACCTGGCCAGAAAGATCTTCTCAATATTACCCTGGCTGGTGGCGGCTCT 2327
DB 181 ACTGAGACCTGGCCAGAAAGATCTTCTCAATATTACCCTGGCTGGTGGCGGCTCT 240
QY 2328 TGTGATGATTTGCCACACACTGGGCGCTTGTTCCTGCTAATTCAGAGGAGATTCAGG 2387
DB 241 TGTGATGATTTGCCACACACTGGGCGCTTGTTCCTGCTAATTCAGAGGAGATTCAGG 300
QY 2388 GCCCAGAGCTGAGACAGATCATCAGATTAACCCAGGCTGGAGAGAGGCCCAAGGCTG 2447
DB 301 GCCCAGAGCTGAGACAGATCATCAGATTAACCCAGGCTGGAGAGAGGCCCAAGGCTG 360
QY 2448 CTGGTGGGCTGACTGCTGGGCGGAGTTGAGGAGAAATGACTGCTTCCACACCACTCA 2507
DB 361 CTGGTGGGCTGACTGCTGGGCGGAGTTGAGGAGAAATGACTGCTTCCACACCACTCA 420
QY 2508 AGTGGGATTAAGTACTGTTTCTTCCACTTAA 2544
DB 421 AGTGGGATTAAGTACTGTTTCTTCCACTTAA 457

RESULT 7
US-10-027-632-141597
; Sequence 141597, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR APPLICATION NUMBER: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-30
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-30
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
```



```

; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 141597
; LENGTH: 808
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-141597

Query Match      17.7% Score 456.6; DB 14; Length 808;
Best Local Similarity 99.8%; Pred. No. 1e-106;
Matches 456; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2068 CTGGCGGGGTTGGTGGCCCACTACTACACCCCTGCTGGGGCTTTTCTTGGAGGCGCTG 2147
DB 1 CTGGCGGGGTTGGTGGCCCACTACTACACCCCTGCTGGGGCTTTTCTTGGAGGCGCTG 60
QY 2148 GTTACAGTGTGGCCGAGGAGATCCCTTCCAGACAGACAGATTGACAAATATGCTTC 2207
DB 61 GTTACAGTGTGGCCGAGGAGATCCCTTCCAGACAGACAGATTGACAAATATGCTTC 120
QY 2208 CAACCTGACAGAGGCTTGTCTTCTCAGACAGAGAGTACCCAGCCAGCCGAGAGAGAC 2267
DB 121 CAACCTGACAGAGGCTTGTCTTCTCAGACAGAGAGTACCCAGCCAGCCGAGAGAGAC 180
QY 2268 ACTGTGACCTGGCCAGAGAAATCTTCTCAATATTAACCCGGCTGGTGGCGGCTCT 2327
DB 181 ACTGTGACCTGGCCAGAGAAATCTTCTCAATATTAACCCGGCTGGTGGCGGCTCT 240
QY 2328 TGTGATGATGATTCGCCACCACTGGGCGCTTTTCCCTAATTCAGAGGAGATTCCAGG 2387
DB 241 TGTGATGATGATTCGCCACCACTGGGCGCTTTTCCCTAATTCAGAGGAGATTCCAGG 300
QY 2388 GCCCAGAGCTGGACAGACATCACAGATTAACCCAGGCTGGAGAGAGGCCCCACGGCTG 2447
DB 301 GCCCAGAGCTGGACAGACATCACAGATTAACCCAGGCTGGAGAGAGGCCCCACGGCTG 360
QY 2448 CTGATGGGGTGTGACCTGGGGGAGTTGGAGGAAATGACCTGCGCTCACCACCAACCCAA 2507
DB 361 CTGATGGGGTGTGACCTGGGGGAGTTGGAGGAAATGACCTGCGCTCACCACCAACCCAA 420
QY 2508 AGTGTGGGATTAACTACTGTTTCTTTCCACTTAAA 2544
DB 421 AGTGTGGGATTAACTACTGTTTCTTTCCACTTAAA 457

RESULT 8
US-09-918-995-30489
; Sequence 30489, Application US/09918995
; Publication No. US2003007623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 30489
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: (1)...(499)
; OTHER INFORMATION: n = A,T,C or G
; US-09-918-995-30489

Query Match      17.3% Score 445.4; DB 11; Length 499;
Best Local Similarity 99.8%; Pred. No. 6.7e-104;
Matches 446; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1169 CCAGCCGAGCTCTGCGGGGCGCCGCCAGATCAAGGCTGTGCTGGAGCTGTGCCCGCTGG 1228
DB 53 CCAGCCGAGCTCTGCGGGGCGCCGCCAGATCAAGGCTGTGCTGGAGCTGTGCCCGCTGG 112
QY 1229 CCGCTCTGTTCTGAGACTGTTTGTGAGAGCCAGCTGTATACCCGACTGCTC 1288
DB 113 CCGCTCTGTTCTGAGACTGTTTGTGAGAGCCAGCTGTATACCCGACTGCTC 172
QY 1289 CTTCCAGGCGCAGCCCTTCACTGTGTGATGTGTGACAACTTTGGGGGAAACATGCTCT 1348
DB 173 CTTCCAGGCGCAGCCCTTCACTGTGTGATGTGTGACAACTTTGGGGGAAACATGCTCT 232
QY 1349 TTTTGGAGCCCTAGAGAGCTGTGACAGAGGCCAGAGCTGCCCTTTCCCACTC 1408
DB 233 TTTTGGAGCCCTAGAGAGCTGTGACAGAGGCCAGAGCTGCCCTTTCCCACTC 292
QY 1409 CACCATGTAGGCAAGGCGATGCCCCCGAGGAGCATCAGCCAGAACGAGTGTCTATTTC 1468
DB 293 CACCATGTAGGCAAGGCGATGCCCCCGAGGAGCATCAGCCAGAACGAGTGTCTATTTC 352
QY 1469 CCTCATGCTGAGCTGGGCTGGCGAAGAGACCCAGTGCAGATTGGCAGCTGGGTGAC 1528
DB 353 CCTCATGCTGAGCTGGGCTGGCGAAGAGACCCAGTGCAGATTGGCAGCTGGGTGAC 412
QY 1529 CAGCTTTGCGCCGCGGATGAGGCTCTCCACCTCGAGCAGAGGGGACGCTGAGAGCT 1588
DB 413 CAGCTTTGCGCCGCGGATGAGGCTCTCCACCTCGAGCAGAGGGGACGCTGAGAGCT 472
QY 1589 ACTGCTCGGAGTGTGTCAACTGCTC 1615
DB 473 ACTGCTCGGAGTGTGTCAACTGCTC 499

RESULT 9
US-10-029-386-9071
; Sequence 9071, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEWICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9071
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO U34879.1
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 0.74
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: NT HIT: U34879.1, EVALUATE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P54802, EVALUATE 2.00e-36
; OTHER INFORMATION: EST_HUMAN HIT: AW205518.1, EVALUATE 0.00e+00
; US-10-029-386-9071

Query Match      16.5% Score 426; DB 13; Length 522;
Best Local Similarity 100.0%; Pred. No. 6.1e-99;

```

Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2119 CTGGCTGGGGCTTTTCTCTGGAGGGGCGCTGGTGAAGAGTGGCCCAAGGGCAATCCCTTTCC 2178  
 Db 1 CTGGCTGGGGCTTTTCTCTGGAGGGGCGCTGGTGAAGAGTGGCCCAAGGGCAATCCCTTTCC 60

QY 2179 AACAGCAGCAGTTTGAACAAAATGTCTTCCAACTGAGAGAGGCTTCCTTCAGCAAGC 2238  
 Db 61 AACAGCAGCAGTTTGAACAAAATGTCTTCCAACTGAGAGAGGCTTCCTTCAGCAAGC 120

QY 2235 AAGAGTACCCCAAGCCGCGGAGAGAGACACTGTGACCTGGCCAAAGAAATCTTCCCA 2298  
 Db 121 AAGAGTACCCCAAGCCGCGGAGAGAGACACTGTGACCTGGCCAAAGAAATCTTCCCA 180

QY 2299 AATATTACCCCGCTGGTGGTGGCGCTCTGTGTGATGATTCGCCACCACTGGGCTTGT 2358  
 Db 181 AATATTACCCCGCTGGTGGTGGCGCTCTGTGTGATGATTCGCCACCACTGGGCTTGT 240

QY 2359 TTTCCGCTAATTCAGAGGAGATTCAGAGGCGCAAGCTGACACATCAAGATTAAC 2418  
 Db 241 TTTCCGCTAATTCAGAGGAGATTCAGAGGCGCAAGCTGACACATCAAGATTAAC 300

QY 2419 CCAGGCTTGGAGAGAGGCCCAAGGCTGTGTGGTCTGACCTGGGGGATTTGAGG 2478  
 Db 301 CCAGGCTTGGAGAGAGGCCCAAGGCTGTGTGGTCTGACCTGGGGGATTTGAGG 360

QY 2479 GAAATGACCTGCTTCCACACCAACCAAGTGTGGATTAAGTACTGTTTCTTCCA 2538  
 Db 361 GAAATGACCTGCTTCCACACCAACCAAGTGTGGATTAAGTACTGTTTCTTCCA 420

QY 2539 CTTAAA 2544  
 Db 421 CTTAAA 426

RESULT 10  
 US-09-918-995-24382  
 ; Sequence 24382; Application US/09918995  
 ; Publication No. US20030073623A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 ; FILE REFERENCE: 20411-756  
 ; CURRENT APPLICATION NUMBER: US/09/918,995  
 ; PRIOR FILING DATE: 2001-07-30  
 ; PRIOR APPLICATION NUMBER: US/09/235,076  
 ; NUMBER OF SEQ ID NOS: 38054  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 24382  
 ; LENGTH: 449  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-918-995-24382

Query Match 15.3%; Score 395; DB 11; Length 449;  
 Best Local Similarity 99.8%; Pred. No. 5e-91;  
 Matches 406; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2150 TGAAGTGTGGCCCAAGGCAATCCCTTTCCAAAGCAAGCTTGAACAAAATGTCTTCCA 2209  
 Db 33 TGAAGTGTGGCCCAAGGCAATCCCTTTCC-ACAGCACAGTTTGAACAAAATGTCTTCCA 91

QY 2210 ACTGAGCAGAGCCTTGTCTCAGACAGAGTACCCAGCCAGCCGAGAGAGACAC 2269  
 Db 92 ACTGAGCAGAGCCTTGTCTCAGACAGAGTACCCAGCCAGCCGAGAGAGACAC 151

QY 2270 TGTGACCTTGGCCAAAGATCTTCTCAATATTAACCCGGGTGGGTGGCGGCTCTTG 2329  
 Db 152 TGTGACCTTGGCCAAAGATCTTCTCAATATTAACCCGGGTGGGTGGCGGCTCTTG 211

QY 2330 GTGATGATTGCGACCACTAGGCTTGTCTTCCGCTAATTCAGGGAGATTTCCAGGCG 2389

Db 212 GTGATGATTGCGACACACTGGGCTTGTCTTCCGCTAATTCAGGGAGATTTCCAGGCG 271  
 QY 2390 CCAGAGCTGGACAGACATCAAGATTAACCAAGCCTTGGAGAGAGGCCCAAGGCTTGT 2449  
 Db 272 CCAGAGCTGGACAGACATCAAGATTAACCAAGCCTTGGAGAGAGGCCCAAGGCTTGT 331

QY 2450 GGTGGGCTTGAACCTTGGGGGATTTGAGGGAATGACCTTGCCTCCACCAACCAAG 2509  
 Db 332 GGTGGGCTTGAACCTTGGGGGATTTGAGGGAATGACCTTGCCTCCACCAACCAAG 391

QY 2510 TGTGGATTTAAGTACTGTTTCTTCCACTTAAAAA 2556  
 Db 392 TGTGGATTTAAGTACTGTTTCTTCCACTTAAAAA 438

RESULT 11  
 US-09-918-995-21830  
 ; Sequence 21830; Application US/09918995  
 ; Publication No. US20030073623A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 ; FILE REFERENCE: 20411-756  
 ; CURRENT APPLICATION NUMBER: US/09/918,995  
 ; PRIOR FILING DATE: 2001-07-30  
 ; PRIOR APPLICATION NUMBER: US/09/235,076  
 ; NUMBER OF SEQ ID NOS: 38054  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 21830  
 ; LENGTH: 468  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(468)  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; US-09-918-995-21830

Query Match 13.9%; Score 358.4; DB 11; Length 468;  
 Best Local Similarity 99.2%; Pred. No. 1.e-81;  
 Matches 359; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2195 CAAAATGTCTTCCAATCTGAGAGAGGCTTGTCTCAGACAGAGATCCCAAGCCA 2254  
 Db 29 CAAAATGTCTTCCAATCTGAGAGAGGCTTGTCTCAGACAGAGATCCCAAGCCA 88

QY 2255 GCCCGAGAGACACTGTGACCTGGCCAAAGATCTTCTCAATATTAACCCGGCTG 2314  
 Db 89 GCCCGAGAGACACTGTGACCTGGCCAAAGATCTTCTCAATATTAACCCGGCTG 148

QY 2315 GTGGCCGCTTGTGTGATGATTGCCACCACTGGGCTTGTCTTCCGCTAATTCAG 2374  
 Db 149 GTGGCCGCTTGTGTGATGATTGCCACCACTGGGCTTGTCTTCCGCTAATTCAG 208

QY 2375 GGCAGATTCCAGGGCCCAAGCTTGAACATCAAGATTAACCAAGGCTTGGAGAG 2434  
 Db 209 GGCAGATTCCAGGGCCCAAGCTTGAACATCAAGATTAACCAAGGCTTGGAGAG 268

QY 2435 GCCCAAGGCTGTGTGTGGGGTGTGACCTGGGGGATTTGAGGGAATGACCTGCCCTC 2494  
 Db 269 GCCCAAGGCTGTGTGTGGGGTGTGACCTGGGGGATTTGAGGGAATGACCTGCCCTC 328

QY 2495 CACCAACCAACCAAGTGTGGATTAAGTACTGTTTCTTCCACTTAAAAA 2554  
 Db 329 CACCAACCAACCAAGTGTGGATTAAGTACTGTTTCTTCCACTTAAAAA 388

QY 2555 AA 2556  
 Db 389 AA 390



QY	1793	TCGGCAGGCACTGAGAGAGCTGCTGCTGACTTGTACTATGAGAGAGGCAGAAAGCCGCTTACT	1852
Db	1665	CCGGCAGAGCACTTCACAACCCGAGCCGGATACCTCTCCCGAGATCAAGAGAGGGGTACGA	1724
QY	1853	GAGCAGAGAGCTGAGCTCCCTCTGTTGAGGGCTGAGAGCGCTCCGTGACTATGAGCTGAGCC	1912
Db	1725	GGCGAAGBACACCCGCCGCTTCGACCCGGCTBACCCGGCGTCTGG---CTGGGCCGTAGTGA	1781
QY	1913	GGCACTGGACGAGGTGCTGGCTAGTACAGCCGCTTCTTGTGGGACAGCTGGCTAGAGCA	1972
Db	1782	CTCTGTGGAGCACTCTCTGGCCACCGACTCCCGTACATCTGCTGGAGCGGTTGGGTGGCGA	1841
QY	1973	GGCCCGAGACAGCGGCACATGAGTGAAGCCGAGACCGGATTTTACAGAGCAGAACACCCGCTA	2032
Db	1842	CGCCCGGGCGGTGGGGGTGCGAGCGCGGCGGACCGGACCGGCTCGGGTACGAGCCCTTCTC	1901
QY	2033	CCAGCTGACTTTGTG-----GGGGCCAGAAAGGCAACATCTGAGCTATGCCAA	2080
Db	1902	GCTGCTGACCTGTGTGGGACACCCCGCGGGGTGCGGACGCGGGGCTGGCGGACTACGCCAA	1961
QY	2081	CAAGCAGCTGGCGGGGTTGTGCGCACTACACCCCTGTGCTGGCGGCTTTTCTTGGGA	2140
Db	1962	CCGCGAGTGGCGGGACTGTCTGGCGGGCTTACCGGCTGGCTGGTGTGACGTACTTCTGC	2021
QY	2141	GGCGCTG 2147	
Db	2022	GGAGCTG 2028	

RESULT 13  
US-10-156-761-1

```

Sequence 1, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIYA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HOSHIKAWA, HIROSHI
APPLICANT: SHIBU, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHISA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURES:
NAME/KEY: misc feature
LOCATION: (4157715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

	Query Match	Similarity	Score	DB	Length
Best Local	961	46.9%	193.4	15	9025608
Matches	961	Conservative	0	Mismatches	996
				Indels	90
				Gaps	9
QY	140	TCCTCTGACCGGGGCGCGGGGCGCGGACGACGACGAGCCCGGAGAGCGGGCGGCGTGC	199		
Db	2453665	TCCTCCCGCGGGGTCCGGGGGCGCGCGCTGCCTGTACC--CGGCTCGGCGGACGG	24537221		
QY	200	GCGCGCTGTGGCCGCGCTGTGGGCGACAGGCCCCGCGGCCGACATTCTCCGTGTGGTGGAA	259		
Db	2453722	AGCGCGCGCGGCGGGGCTCGCGCGCGACGAGCGCGCGCGCGGCTCTCTCCCGCTCACTGGCG	2453781		

OY	260	GCGGCTCTGTGGCTGCCAAGCCGGGCTTTGGACACTTCAACGCTTGGGGGCGGGCGGGCGGC	319
Dd	2453782	GCACCTCAGCTTCCTCGCACGGCGGGCGGACCGGACAAGTTTCAGAGTGTCCGGCCGTAACGGG	2453841
OY	320	GCGGCTCGGGATGGCGGCGGCTCCACCGAGCGTAGGGCGCGCGCGCGGCGTGCACCGCTACCT	379
Dd	2453842	CCGGGTACCGGTACCGGGGGGGGACACCGGGGACCAGGCTCAGGGGGCTGAACGTGACTT	2453901
OY	380	GCGGCACTTCTGTGAGCTGCCACGTTGGCTGTGTCCGGCTCTCACTGCGCTGCGCGGGCC	439
Dd	2453902	GAGGAACATCGCGGACCCCGACATCACTGGGGGGGCGAGCACTCCGCTGCCCCGGC	2453961
OY	440	ACTCCAGCGCGTGGCGGGGAGCTGACCCGAGCGCACGCCCAACAGATTACCGCTATTACCA	499
Dd	2453962	ACTGCCCCGGCTCGCGCGGACAGTAGACCCGGCGCGGAAAGTCCCGACCGCTTGGCCCT	2454021
OY	500	GAATGTGTGACCGCAAAAGTACTTCTGTTGTGTGGAGCTTGGGCCCGCTGGAGCGAGA	559
Dd	2454022	CACGACACGACACGACCGGTACAACGGGCGCGTACACAGCTGACGTAACGTAGGGAGGTGA	2454081
OY	560	GATGACCTGATGGCGGTGATGGCATCACTGGACCTGAGCGTGGAGCGGCGAGGAGCG	619
Dd	2454082	GCTGACGTACTGGCCTTGCAAGCTTACACGAAGTCTGTCCAACGGGCGCGGACGC	2454141
OY	620	CATCTGGAGCGGATGTACTTGCCCTTGGGCTGACCACGACGACAGATCAATGATTTCTT	679
Dd	2454142	GCTCCACACCGGGGTGTTCCAGGAGTTGGGATACCGACGAGAGAGCTCGGAAGTGAT	2454201
OY	680	TACTGTGTCGTGCTCTCTGGGCTTGGGGGCGAATGGCGAACCTGCACACTGGAGATGGCCC	739
Dd	2454202	TCCGGGGCGGGCCATCAGCCCTGTGTGGCTGACAGAACCTTCTGCTTCCCGCACCC	2454261
OY	740	CCTGCCCCCTCTCTGGACATCAAGCAGCTTACCTTGACACACCGGATCTCTGACACGAT	799
Dd	2454262	CGTCTCAACAAAGTCTCTGCACCGCGGGGCGGGCGCTGGCGGGGATTCGCAACCGGCT	2454321
OY	800	GCGCTCTTTGGGCAATGACCCCAAGTGTGCTGCAATTTGGGGGAGATTTCCGAGGCTGT	859
Dd	2454322	GCGGAGCTGGGACATGACACCGGGTGTCCCGGCTATTGGGACAGTTTCGCGGGCTT	2454381
OY	860	CACCAAGGATGTTCCTCAGGTCAATGTACGAAATGGGAGTGGGGGCACATTAACTG	919
Dd	2454382	CGCCGACCGGACCGCGGGTGGGACACCGTCCCGGAGGACCTGANTGGGCTTCCGACG	2454441
OY	920	TTCTACTCTGTCTCTTCTCTTGTGGGCTCGGAAAGCCCATATTTCCCATCATCGGAG	979
Dd	2454442	GCCGACATGG-----CTGACCGCGGACCGAACATTCAACGGGTGGCGCGGC	2454492
OY	980	CCTCTTCTGGAGAGCGATCAAAAGATTGGCACAGACCAATCATATGAGGGCGGACAC	1039
Dd	2454493	CTTCAACCGGATCCAGACGAGATGTTGGGGGGGCTCGACCGCTTCAAGATGACCT	2454552
OY	1040	TTTTCAATGATGACGCCACTTTCCTAGAGCCCTCTTACTTTCGCGAGCCACCACTGC	1099
Dd	2454553	GCTGACGAGGGCGGACGCCCGCGGTGACGTGTC-----CGTGGGCAACCGGGAAGGG	2454606
OY	1100	CGTCTATGAGGCGCATGACTGTAGATGTACTGAGGCTGTGTGGGTGTGCTCCAGGCTGAGCT	1159
Dd	2454607	CGTCAAGAGGCGGTGGCGGGCGGACCCGGGCGCGGTATGGGTGATCTCGGGCTGGC-	2454665
OY	1160	CTTCACAGCACGAGCGGAGTTCTGGGGGCGCGCCAGATCAGGAGCTGTGCTGGAGCTGT	1219
Dd	2454666	-----AGCAACAACCGGCGAGGCGATTCGCGACGCGGT	2454699
OY	1220	GCCCCGTGCGGCTCTCTGTTCTTGACCTGTTCCTGAGGCGCAGCTGTGTATAC---	1276
Dd	2454700	GGACAGAGACCGGATGCTCGTGTGAGACGCGCTGTGACCGCTTCCGAAAGTCAACGA	2454759
OY	1277	CCGCACTGCTCTCTTCCAGGGCGAGCCCTTATCTGGTGCATGCTGCACAACATTGGGGG	1336
Dd	2454760	CCGGAGGCGGACGTGACACCGGACCCGTAACGCTTGGCTCGATCTGGAACCTTGGGGG	2454819
OY	1337	AACCACTGCTTTTGGAGCCTTAGAGGCTGTGACGAGGCGCCAGAGCTGCCGCT	1396

```

Db 2454820 CCAATGACCTGCGGCGGAAACCTCCGACTGAGCTTGATGACGAGCGGTGCGCAC 2454879
Qy 1397 CTTCCCACTCCACCATGATAGGACAGGACATGAGGACCATCAGCAGAACGA 1456
Db 2454880 CAGGCGGGGACGACCTCGCGGGGTGTGTCTGTCTCCGAGGCGCGGACAACTCC 2454939
Qy 1457 AGTGTCTATTTCCTCATAGCTGAGCTGGGCTGGGAAAGAACCAAGTCCAGATTGGC 1516
Db 2454940 GAGCGCTTCGCGCTGTTCCTGAACTCGCTGGGCGAGGACACTTGG--GACTGCG 2454996
Qy 1517 AGCGTGGGTGACCAAGCTTTCGCGCGGCGGTATGGGGTCTCCACCGGACGACGAGGCG 1576
Db 2454997 CGCATGATTCGGCGGTGGGCGCGCTCGCGCTACGGGGAGACCATTCACATGCGAGGC 2455056
Qy 1577 AGCGTGAAGGTACTGCTCCGAGAGTGTATACATGCTCCGAGGAGGCGCTGACAGGCGCA 1636
Db 2455057 CGCGTGGACATCTCGCGCGGACCGGTACGAGCACGCGCGCGCACTGTGAGCGA 2455116
Qy 1637 CAATGTAGCCCGCTGTGACAGGCGCGCTCCCTACAGATGAATACA----- 1683
Db 2455117 GGGCGGAGCGGCTGTTCGGGCGCGCGCTCCCTCGCGGCGACGAGGCGCTTCTG 2455176
Qy 1684 -----GCATCTGTATCAACCGATCTGATGTGTTTGAAGCTGCGCTGCTGCT 1732
Db 2455177 GTCCGCGAAGCGCTCCGTTACCGGCGCGAGAGTTCGAAACGGCGCTTGGCGAGCTCCT 2455236
Qy 1733 CACATCTGCTCCCTCCCTGCGCACACCGCCCGCTTCCGCTTACGACTCTGTGACCTTAC 1792
Db 2455237 GAAAGTGGCGCGCGGCGGTGCGGCGCTCGTCCGCGTACCGCGCACTGCTGCACTGTGC 2455296
Qy 1793 TCGGACGAGCAGTGCAGAGCTGTGTGACTGTATGAGGAGGACAAAGCGCTTACT 1852
Db 2455297 CGCGAGGACATCTTCAACCGACCGGCTACTCTCCCGAGATCAGAGAGCGGTACGA 2455356
Qy 1853 GAGCAAGAGCTGGCTTCCCTGTTGAGGCGTGAAGCGCTGCGCTATGAGCTGTGCC 1912
Db 2455357 GAGCAAGAGCACCGGCGCTTGAACCGGCTGACCGGCGCTGTG--CTGGCGCTGATGGA 2455413
Qy 1913 GGCACGTGACAGAGTGTGCGGTAGTACAGACCGCTTCTTGGGCGAGCTGAGAGCA 1972
Db 2455414 CCTGTGAGAGCACTCTGCGCACCGACTCCGCTATGTGTGGCGCGTGTGGTGGCGGA 2455473
Qy 1973 GCGCCGACAGCGGAGTCAAGTGAAGCGGACGATTTCTACAGACGAAACAGCGCTA 2032
Db 2455474 CGCGCGGCGGTGGGTGAGAGCGGAGCGGAGCGGCTCGGTACGACGCGCTCTC 2455533
Qy 2033 CCAAGTGACTTTGTG-----GGGCGAGAAAGGCAACATCTTGAATATGCCAA 2080
Db 2455534 GCTGTGACCGTGTGGGACCGCGCGGCGGTGCGGACCGCGGCTGCGCACTACGCGCAA 2455593
Qy 2081 CAAGCAGCTGGCGGGTGTGTGCACTATACACCGCTGCGGCGCTTTCTGGA 2140
Db 2455594 CGCGAGTGGGCGGACGTGTGCGGCGGTATCCGCGTGTGCGTGTGACGTACTTCC 2455653
Qy 2141 GCGCGCTG 2147
Db 2455654 GAGCGT 2455660

```

```

RESULT 14
US-10-156-761-5965
; Sequence 5965, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: SHIBATA, TADAYOSHI
; APPLICANT: SHIBATA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRU
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

```

```

; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5965
; LENGTH: 3114
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3114)
; US-10-156-761-5965

Query Match 6.0%; Score 155; DB 15; Length 3114;
Best Local Similarity 51.7%; Pred. No. 1,6e-29;
Matches 353; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy 178 CCGGAGGCGGCGCGCTGCGGCGCTGTGACCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGG 237
Db 101 CCGGAGGCGGCGCGCTGCGGCGCTGTGACCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGG 160
Qy 238 CGGACCTTCGCGGTGTGAGCGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 297
Db 161 CCGGAGATCGGACATGTTCCGCTTATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 220
Qy 298 GCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 357
Db 221 TCGGAGTCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 280
Qy 358 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417
Db 281 TACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 340
Qy 418 CTAAGTGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 477
Db 341 GCGAGTGAACCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 400
Qy 478 CCAACAGGTACCGGTATTCAGGAATGTGTGACCGGAAAGCTTCTTGTGTGTGTGTGTGTGT 537
Db 401 CCGTGTCCGCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTT 460
Qy 538 ACTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 597
Db 461 ACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 520
Qy 598 TGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 657
Db 521 TGGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 580
Qy 658 AGGCAAGATCAATAGTCTTTTACGTGCTGCGCTTCTGCGCGGCGGCGGCGGCGGCGGCGG 717
Db 581 ACAAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 640
Qy 718 ACCTGACACCTGGAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777
Db 641 ACCTTCGCGGATACCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 700
Qy 778 AGGACCGGATCTGGAACGAGATGCGCTCTTGTGAGTACCGGCGGCGGCGGCGGCGGCGG 837
Db 701 GCGCGCGGATCTGCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 760
Qy 838 CCGGCGGATGTTCCCGAGGCTGTG 860
Db 761 ACGGCGACGTCCCGAAGGCTTC 783

```

```

RESULT 15
US-10-156-761-1/c
; Sequence 1, Application US/10156761

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2004, 13:32:04 ; Search time 54 Seconds  
(without alignments)

2183.960 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939

Sequence: 1 MEAVAVAAAVGVLLLAGAGS.....VDIAKIFLKYIPGWAGSW 743

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: A\_Geneseq\_19Jun03:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3939	100.0	743	18	AAW18017
2	1407	35.7	690	22	ABE68119
3	147.5	3.7	101	23	ABP32743
4	125.5	3.2	497	22	AAE56171
5	121	3.1	4342	22	AAU53611
6	121	3.1	4472	18	AAW22601
7	120	3.0	3729	18	AAW22603
8	118.5	3.0	871	22	AAE30570
9	118.5	3.0	871	23	AAE16274

10	118.5	3.0	1351	22	AAE30567	Amino acid sequenc
11	118.5	3.0	1610	22	AAE30568	A full length huma
12	118.5	3.0	1618	22	AAE85504	Human protein kin
13	118.5	3.0	1665	23	AAE24151	Human kinase (PKI
14	118.5	3.0	1665	23	AAE15372	Human tyrosin lig
15	118.5	3.0	2566	22	AAE30569	A splice variant o
16	118.5	3.0	2630	24	ABG76186	Human serine/threo
17	118.5	3.0	7968	24	ABG76187	Human serine/threo
18	117.5	3.0	497	16	AAE72376	Human auxillary cy
19	117.5	3.0	497	17	AAE93183	Human auxillary cy
20	117.5	3.0	497	17	AAE93184	Human auxillary cy
21	117.5	3.0	497	17	AAE93185	Human auxillary cy
22	117.5	3.0	497	17	AAE93186	Human auxillary cy
23	117.5	3.0	497	17	AAE93187	Human auxillary cy
24	117.5	3.0	497	23	ABE09563	Cytochrome P450Iid
25	117.5	3.0	497	23	ABE09564	Human CYP2D6 prote
26	113	2.9	7257	21	AAE58576	Human drug-metabol
27	111.5	2.8	484	21	AAE45186	Sorangium cellulos
28	111.5	2.8	497	16	AAE72375	Arabidopsis thalia
29	111.5	2.8	497	16	AAE72376	Human auxillary cy
30	111.5	2.8	497	17	AAE93182	Human auxillary cy
31	111.5	2.8	497	17	AAE93183	Human auxillary cy
32	109.5	2.8	1377	22	AAU34791	Human cytochrome p
33	109.5	2.8	1377	22	AAE98997	E. coli cellular p
34	107.5	2.7	1621	19	AAE60162	E. coli growth and
35	106	2.7	308	22	AAE92974	Arabidopsis thalia
36	106	2.7	310	23	AAE51006	Corynebacterium gl
37	106	2.7	1005	22	ABG11447	Novel human diago
38	105	2.7	1203	21	AAE81572	Novel human endogen
39	105	2.7	4150	21	AAE92707	Mus dunni endogeno
40	104.5	2.7	1622	19	AAE60163	S. antibioticus 8,
41	104	2.6	581	21	AAE84322	Arabidopsis thalia
42	104	2.6	581	21	AAE85701	A human cardiovascular
43	104	2.6	588	22	AAE59161	Novel protein kin
44	104	2.6	588	23	AAU11406	Adeno associated v
45	104	2.6	654	22	AAU29036	Adeno associated v

## ALIGNMENTS

RESULT 1	AAW18017	AAW18017 standard; Protein: 743 AA.
ID	AAW18017;	
AC	AAW18017;	
XX	20-AUG-1997 (first entry)	
DT		
XX	Human alpha-N-acetylglucosaminidase.	
DE		
XX	Alpha-N-acetylglucosaminidase; mucopolysaccharidosis type IIIB;	
KW	gene therapy; enzyme replacement therapy; diagnosis.	
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Peptide	1..23
FT	Protein	/label= Stg_peptide
FT		22..743
FT		/label= Mat_protein
FT	Modified-site	261
FT		/note= "potential N-glycosylation site"
FT	Modified-site	272
FT		/note= "potential N-glycosylation site"
FT	Modified-site	435
FT		/note= "potential N-glycosylation site"
FT	Modified-site	503
FT		/note= "potential N-glycosylation site"
FT	Modified-site	513
FT		/note= "potential N-glycosylation site"
FT	Modified-site	526
FT		/note= "potential N-glycosylation site"



FT Modified-site 532 /note= "potential N-glycosylation site"

XX MO9719177-A1.

XX 29-MAY-1997.

XX 22-NOV-1996; 96MO-AU00747.

XX 23-NOV-1995; 95AU-0006748.

XX (WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.

XX Anson DS, Blanch L, Hopwood JT, Scott H, Weber B;

XX WPI; 1997-298114/27.

XX N-PSDB; AAT67163, AAT67164.

PT Nucleic acid encoding mammalian alpha-N-acetylglucosaminidase -

PT used for the diagnosis and treatment of mucopolysaccharidosis type

PT IIB, also used in gene therapy

PS Claim 11, Page 46-50; 79pp; English.

XX Human alpha-N-acetylglucosaminidase (AA18017) can hydrolyse the

XX terminal alpha-N-acetylglucosamine residues at the non-reducing

XX terminus of fragments of heparan sulphate and heparin. Its amino

XX acid sequence was deduced from a human peripheral blood leukocyte

XX cDNA clone (AAT67163). Recombinant enzyme can be expressed in host

XX (pref. CHO) cells and may be expressed as a fusion to e.g. an

XX enzyme, reporter molecule, purification tag and/or signal sequence.

XX It can be used to treat alpha-N-acetylglucosaminidase deficiency,

XX for example in patients suffering from mucopolysaccharidosis type

XX IIB. Administration is by oral, i.v., i.p., enzyme replacement

XX therapy, gene therapy or other routes.

XX Sequence 743 AA;

SO Query Match 100.0%; Score 3939; DB 18; Length 743;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAAVAAAVGVLLLAGAGAGAGDEARBAAVATVRLIGGPADFSVSEVERLAAKP 60

DB 1 MEAAVAAAVGVLLLAGAGAGAGDEARBAAVATVRLIGGPADFSVSEVERLAAKP 60

QY 61 GLDTTSLGGGGAARVRVRSSTGVAAAAGLHRYLRDPCGGHVAMSGQLPRPLPAVPG 120

DB 61 GLDTTSLGGGGAARVRVRSSTGVAAAAGLHRYLRDPCGGHVAMSGQLPRPLPAVPG 120

QY 121 LTEATPRRYRYQNVCTOSYFVWMDMARWEREIDMWALNGINIALAMSGOBAIMQRYL 180

DB 121 LTEATPRRYRYQNVCTOSYFVWMDMARWEREIDMWALNGINIALAMSGOBAIMQRYL 180

QY 181 ALGLTQAEINEFTGPAFLAMSGMNLHTWDGFLPSPWHIKOYLQHRVLDQMRSGMT 240

DB 181 ALGLTQAEINEFTGPAFLAMSGMNLHTWDGFLPSPWHIKOYLQHRVLDQMRSGMT 240

QY 241 VLPAFGHVEATRVFPQNVNTKMGSGHFNCSYCSFLIAPEDIFIIISLFRELI 300

DB 241 VLPAFGHVEATRVFPQNVNTKMGSGHFNCSYCSFLIAPEDIFIIISLFRELI 300

QY 301 KERGTTHIYGADTFNEMQPPSPSYLAATTAIVTEAMTAVDTBAVWLQGMFLQHPQ 360

DB 301 KERGTTHIYGADTFNEMQPPSPSYLAATTAIVTEAMTAVDTBAVWLQGMFLQHPQ 360

QY 361 WGAQIRAVLGAAPRGRLVLDLFASSQPYRTTASFOQPFVWCMHLHFGNHLFGAL 420

DB 361 WGAQIRAVLGAAPRGRLVLDLFASSQPYRTTASFOQPFVWCMHLHFGNHLFGAL 420

QY 421 EAVNGGEARLFPNSTMVGTGMAPGISQNEVYVSLMELGKRDVPVDLAAWTSFAA 480

DB 421 EAVNGGEARLFPNSTMVGTGMAPGISQNEVYVSLMELGKRDVPVDLAAWTSFAA 480

QY 481 RRYGVSHPDAGAAWRLLLRSVYVNSGEGACRHNRSPLVRPPSIQMTSIWYNSDVEAW 540

DB 481 RRYGVSHPDAGAAWRLLLRSVYVNSGEGACRHNRSPLVRPPSIQMTSIWYNSDVEAW 540

QY 541 RLILTSAPSLATSPAFRYDLDTTRQAVQELVSLYTEAASATLSKLAULLAAGVTLAY 600

DB 541 RLILTSAPSLATSPAFRYDLDTTRQAVQELVSLYTEAASATLSKLAULLAAGVTLAY 600

QY 601 ELPLALDEVILASDRFLLGSMLEQARAAVSEADPEQNSRYQLTMGPEGNIIDYAN 660

DB 601 ELPLALDEVILASDRFLLGSMLEQARAAVSEADPEQNSRYQLTMGPEGNIIDYAN 660

QY 661 KQLAGLVANYTTPRWRFLBALVDVSAQGIPOOHQPKVNFQLECAFVLSKORYPQPR 720

DB 661 KQLAGLVANYTTPRWRFLBALVDVSAQGIPOOHQPKVNFQLECAFVLSKORYPQPR 720

QY 721 GDTVDLAKKIFLKYPGVMVAGSW 743

DB 721 GDTVDLAKKIFLKYPGVMVAGSW 743

RESULT 2

ID ABB68119 standard; Protein; 690 AA.

XX ABB68119;

AC ABB68119;

XX 26-MAR-2002 (first entry)

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 31149.

DE Drosophila melanogaster polypeptide SEQ ID NO 31149.

XX Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical.

XX Drosophila melanogaster.

OS Drosophila melanogaster.

XX WO200171042-A2.

PN WO200171042-A2.

XX 27-SEP-2001.

PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

PR 11-JUL-2000; 2000US-0614150.

XX (PEXE ) PE CORP NY.

PA (PEXE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

XX WPI; 2001-656860/75.

DR N-PSDB; ABL12222.

XX N-PSDB; ABL12222.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

PS Disclosure; SEQ ID NO 31149; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA

XX sequences (AB101840-AB161175) and the encoded proteins

XX (AB55737-AB572072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 690 AA;

Query Match	35.7%;	Score 1407;	DB 22	Length 690;
Best Local Similarity	43.3%;	Pred No. 1.9e-118;		
Matches 290;	Conservative 102;	Mismatches 255;	Indels 22;	Gaps 11
QY	74	RVEVRSSTGYAAAAAHRYLRIDFCGCHVAMSGSQLRPR--PLPVBGELTEATPNRYR	131	
Db	6	RILLMGDGVSVCALHLYLKYLVINDVDWFMKRIEPTNLQLPNVTIEKSAASP--IIY	63	
QY	132	YQNVCTQSYFVFWMDMAAREBELDMWALNGINLALMSGCAIMQRYVALGLQIAEINE	191	
Db	64	HQNVCTWSISFAAWGIEQWRRLDMWALNGISLTIA-PVQALVWKYTDMGRLMEIDE	122	
QY	192	FFTGPAFLAWKNGINLHTWDGCLPSPWHIKQLYQHRVLDQMSRFGVTPYLPAGHVE	251	
Db	123	HLGAPFAQAGQNRGINRGMAQPLTPAMRRYQLLQDEITIQNRMLGMSVALPAPAGHYR	182	
QY	252	AATRVFPQVNVTKMGSGHFNCSYSGSFLIABEDPIFIQSLFLRELKEFGDTIYGA	311	
Db	183	ALAKRLPESTEMVQKWNQFPDRYCCGLFVEFTEMLPEKIASRLNHIIRKYSNHIFFC	242	
QY	312	DTENEMQPSSEPSYLAATTAAYVEAMTAVDTEAWVLQGLVFCQHPQFGPAQIRAVIG	371	
Db	243	DPENELEPPAKETMYRSTRAALIESMGIDPQALWLLQGMFPYAN-FITTMARAFUL	301	
QY	372	AVPRGRLLVLDLFAESQPVYTRTASFPQGPFTWMLHNGNGHGLFGALFAVNGEPEAR	431	
Db	302	AAERGRILLVDLQSEQFPQYELTRSYFGQPFIMCMLHFGGLGLMFGSALKINSIEEAR	361	
QY	432	LPENSTNVQGMAPESIGSNEVYVSLMELGMRQDPVLDLAAWTSPARAYGVSHPDAG	491	
Db	362	RLPNSSISVGTIGTIPBEIGQNYVMYFTLERGNSNTSL-DLDSMTTNSHSRYGVKDERLE	420	
QY	492	AAARLLRSYYNCSG-EACRGHNRSPLYRPELQNTSIVYNSRDVEANRLILTSAPSL	550	
Db	421	QAWLLKNSYVSFRGLQKRG--QYVWTRRPSFNQEPPTYNAGAVLDAMHLLLTFRAT	478	
QY	551	ATS----PAFRYLLDLTRQAVQELVSLYEABARATYISKELASLLAAGVALAYELLPAL	606	
Db	479	PLEDNRYEIYEHLDVITDFQIISADQIYINLRSAYKRRQVSRF----EFLISVKLLKLF	534	
QY	607	DE---VLASDSRFLLSWLEQAPAAAVASHAEADFEQORSRQLTLMPESGNIIDYANKOL	663	
Db	535	DMELLILASRNLLQGNWLOQAKKQAPNPGQQRNFEFARNQITAMKPDQIILDYACKQW	594	
QY	664	AGLVANYTTPKRWLFLELVALDSVAGQIPPIQCHQPKNV-FGLEQAFVLSKQRPSPQRPD	722	
Db	595	SGVSDYDYRRMRLLFLEDVTVALHGRPFNGTAFKLKYSHEIELFENKQDVYPVTPYVG	654	
QY	723	TVDLAKKIF 731		
Db	655	TWLSQDIF 663		
RESULT 3				
ABP32743				
ID	ABP32743	standard;	Protein;	101 AA.
XX	AC	ABP32743;		
XX	DT	08-JUL-2002	(first entry)	
DE	XX	Human ORF1716	protein, SEQ ID NO:3432.	
XX	XX	Human; ORF; open reading frame; ORFX; drug screening; diagnosis;		
XX	XX	disease monitoring; cytokine; cell proliferation; cell differentiation;		
XX	XX	immune modulation; haematopoiesis regulation; tissue growth;		
XX	XX	angiogenesis; actinin; inhibin; chemotactic; chemokinetic; haemostatic;		
XX	XX	chromolytic; tumour inhibition; bodily characteristics; fertility;		
XX	XX	behaviour; cancer; proliferative disorder; neurological disorder;		
XX	XX	cardiovascular disease; immune system disorder; organ transplantation;		
XX	XX	tissue growth disorder; tissue regeneration disorder; diabetes mellitus;		
XX	XX	hypochloridism; cholesterol ester storage disease; infection; vulnerability;		
XX	XX	vasoconstrict; antidiabetic; antidiabetic; cytostatic; nootropic;		

KW neuroproctective; antihaemorrhagic; anticonvulsant; thrombolytic;  
XN cardiant; hypotensive; antihypertoid; antifibrinolytic; immunomodulator;  
XM dermatological; analgesic; virucide; antibacterial; fungicide.  
OS Homo sapiens.

PY WO200190366-A2.

XV  
PN  
PM 29-NOV-2001.

PP 24-MAY-2001; 2001WO-US17076.

PR 24-MAY-2000; 2000US-206690P.

RK (CURA-) CURAGEN CORP.

SF Leach MD, Shinkets RA;

TU WPI; 2002-106200/14.

VZ N-PDSB; AHN76769.

XX Novel human polypeptides and polymucleotides useful for diagnosing,  
YY preventing and treating cardiovascular disease, neurodegenerative,  
ZZ hyperplastic disorders and disorders related to organ transplantation

AA Claim 10; Page 118; 2508pp; English.

AB Sequences ABP31028-ABP35561 represent 4534 novel human proteins designated ORFX (open reading frame) 1-4534, and sequences AHN75054-ABN79587 represent cDNAs encoding them. The invention also encompasses polympeptides at least 80% identical to the ORFI-ORFA434 (collectively referred to as ORFX) proteins, polymucleotides at least 85% identical to the ORFX nucleic acid sequences, vectors and host cells comprising ORFX polymucleotides, the recombinant production of ORFX proteins, antibodies specific for ORFX proteins, methods of detecting ORFX polymucleotides and polypeptides, methods of screening for modulators of ORFX expression or activity, and methods of screening individuals for a predisposition to an ORFX-associated disorder. The ORFX proteins of the invention have a wide range of biological activities, such as cytokine, cell proliferation, cell differentiation, immune modulation, haematopoiesis regulation, tissue growth, angiogenesis, activin or inhibin activity, chemotactic/chemokinetic activity, haemostatic activity, tumour inhibition activity, receptor/ligand, anti-inflammatory activity, tumour inhibition activity, and antimetastatic activity, and may also be involved in the determination of bodily characteristics, fertility and behaviour. ORFX proteins, nucleic acids and antibodies may be used in the treatment of cancers, other proliferative disorders such as psoriasis and benign tumours, neurological disorders such as epilepsy and Alzheimer's disease, cardiovascular diseases, immune system disorders, disorders related to organ transplantation, disorders of tissue growth and regeneration, diseases such as diabetes mellitus, hypothroidism, and cholesterol ester storage disease, and infectious diseases caused by viral, bacterial, fungal and other pathogens. ORFX nucleic acids may also be used as a source of primers and probes, in the detection of ORFX genomic sequences or transcripts, in the identification and cloning of homologous sequences, in genetic diagnosis, and in forensic biology. The ORFX nucleic acids may additionally be used to produce transgenic animals which may be useful for studying the function and/or activity of ORFX protein, and in drug screening. The ORFX proteins may also be used as immunogens to generate specific antibodies, which are useful in the diagnosis, treatment and monitoring of ORFX-associated diseases.

BQ Sequence 101 AA;

CQ Query Match 3.7%; Score 147.5; DB 23; Length 101;

DQ Basic Local Similarity 35.6%, Pred.No. 2.7e+05;

EQ Matches 31; Conservative 18; Mismatches 33; Indels 5; Gaps 1;

FQ GSMTLEQRRAAVSEADDFYNRGRYLDTM-----GPGGNITLDYPANKQLAGLVANYTP 673  
|||||:::|||||:|||||:  
I GPWFESKSALATKEKGQVFMRNRGVTVMYNDTETFGSLTHFYANFWSGLLRSYLP 60

QY 674 RWRFLFALVDVSAQGFPOOHDPDKN 700  
 DB 61 RASKYFAVLTRRLQNRSPQLEWRKD 87

RESULT 4  
 AA05171  
 ID AA05171 standard; Protein; 497 AA.  
 AC AA05171;  
 XX  
 XX 12-SEP-2001 (first entry)  
 DT  
 DE Human drug metabolising enzyme (DME-2) protein.  
 XX  
 XX Human; drug metabolising enzyme; DME-2; immunosuppressive; gene therapy;  
 KW cytotoxic; autoimmune disorder; inflammatory disorder; atherosclerosis;  
 KW osteoporosis; eye disorder; hepatic tumour; Addison's disease; cretinism;  
 KW rheumatoid arthritis; acquired immune deficiency syndrome; AIDS; anaemia;  
 KW developmental disorder; endocrine disorder; iritis; acromegaly; epilepsy;  
 KW thyrotoxicosis; pancreatic disorder; diabetes mellitus; obesity; adenoma;  
 KW gastrointestinal disorder; nodular hyperplasia; conjunctivitis; glaucoma;  
 KW actinic keratosis; metabolic disorder; dysphagia; anorexia; carcinoma;  
 XX cell proliferative disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..25  
 FT /label= Signal\_peptide  
 FT Protein 26..497  
 FT /note= "Mature drug metabolising enzyme (DME-2) protein"  
 FT Domain 1..18  
 FT /label= Transmembrane\_domain  
 XX  
 XX WO200151638-A2.  
 XX  
 XX 19-JUL-2001.  
 PD  
 PF 12-JAN-2001; 2001WO-US01174.  
 XX  
 XX 14-JAN-2000; 2000US-0176139.  
 PR 21-JAN-2000; 2000US-0177443.  
 PR 28-JAN-2000; 2000US-0178574.  
 XX  
 XX (INCYTE GENOMICS INC.  
 PA  
 PI Yang J, Baughn ME, Burford N, Au-Young J, Lu DM, Reddy R;  
 PI Ring HZ, Hillman JL, Yue H, Azimzal Y, Yao MG, Gandhi AR;  
 PI Nguyen DB, Tang YT, Lai P, Bandman O;  
 XX  
 DR WPI: 2001-425874/45.  
 DR N-PSDB; AAD09937.  
 XX  
 XX Drug metabolizing enzymes and encoding polynucleotides, useful for  
 PT diagnosing, treating and/or preventing autoimmune, inflammatory, cell  
 PT proliferative, developmental, endocrine, eye, metabolic, and  
 PT gastrointestinal disorders -  
 XX  
 PS Claim 1; Page 136-137; 133pp; English.  
 XX  
 CC The present sequence is human drug metabolising enzyme (DME-2) protein.  
 CC Human DME and its nucleic acid molecule are useful for the diagnosis,  
 CC treatment and prevention of disorders associated with increased or  
 CC decreased expression of DME. Examples of such disorders include,  
 CC autoimmune/inflammatory disorder such as acquired immune deficiency  
 CC syndrome (AIDS), rheumatoid arthritis, osteoporosis; cell proliferative  
 CC disorder such as actinic keratosis, atherosclerosis; developmental  
 CC disorder such as epilepsy, anaemia, endocrine disorder such as  
 CC acromegaly, cretinism, thyrotoxicosis; pancreatic disorder such as  
 CC diabetes mellitus; eye disorder such as conjunctivitis, glaucoma, iritis;  
 CC metabolic disorder such as Addison's disease, obesity; gastrointestinal  
 CC disorder such as anorexia, dysphagia and hepatic tumours including

CC nodular hyperplasia, adenomas and carcinomas. DME DNA is useful for  
 CC creating 'knockin' humanised animals (pigs) or transgenic animals (mice  
 CC or rats) to model human disease. DME DNA is also in useful is gene  
 CC therapy. DME and its immunogenic fragments are useful for screening  
 CC libraries of compounds in several drug screening assays.  
 CC  
 XX  
 SQ Sequence 497 AA:  
 Query Match 3.2%; Score 125.5; DB 22; Length 497;  
 Best Local Similarity 21.9%; Pred. No. 0.033;  
 Matches 93; Conservative 53; Mismatches 143; Indels 135; Gaps 21;  
 QY 345 AVNLL-----GGLFQHQPFGWPAQIRAVLGVPRGRLLVLDLFAESQPVYTRRTAS 396  
 DB 14 AIFLLVLDLMDHRRQRRAARVPP--GDLPLPGL-----GNLLHYD----- 50  
 QY 397 FQCGPFTWCLNHNFGN-----HQLFGALBA-VNGGEAARL--FPNSTWV 439  
 DB 51 FQNTFYCFQGLRRRPFDSVSLQAWTPVYVWGLAAVRALVTHGEDTADRPVPITQIL 110  
 QY 440 GTMAPEGI-----SQNEVYSLMAELGMRDPPVDLAWTSTAARKYGVSHPD 489  
 DB 111 GFPRSQGVFLARYGPAMREQRFSVSTLRNLGLK--KSLBQWTEEA----- 158  
 QY 490 AGAAMRLLRVSYNCGEACRGHNRSPILVRPSLQWNTSITWNRSPFEAMRLLLTSAPS 549  
 DB 159 -----CLCAAFANHSGRPF--RNGLLDKAV-----SNVI-----AS 188  
 QY 550 LATSPAERYD-----LLDLTROAVGELVSLYEARSAYLSKEIASLIRAGCVLAE-- 601  
 DB 189 LTCGRFFEYDDPFRLLDLPLACGLKEBSGFLREVINAVPVPHIPAL--AGVIRFFORA 246  
 QY 602 LPLALDEVL-----ASDSRFLIGSWLBOARAAVSEADPYEONSRYQLTLMGPE 652  
 DB 247 FLTQDLBELTHERMTWDPAPQPPDLTFAFLAK--KEKAKSPSSSFNDENLRIV----- 299  
 QY 653 GNILDYANKQLAGLVANYTYTPRWRFLFLEALVDVSAQGFPOOHDPDKNVFQLEQAFVLAK 712  
 DB 300 GNLF-----LAGWVTSSTLLAWALLMLTHPVGCRV--QOEIDEVIGQVHRPMDAQ 350  
 QY 713 QRYP 716  
 DB 351 AHMP 354

RESULT 5  
 AAU33611  
 ID AAU33611 standard; Protein; 4342 AA.  
 XX  
 XX AAU33611;  
 AC  
 XX  
 XX 14-FEB-2002 (first entry)  
 DT  
 DE Pseudomonas aeruginosa cellular proliferation protein #55.  
 XX  
 XX Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Pseudomonas aeruginosa.  
 XX  
 PN WO200170955-A2.  
 XX  
 XX 27-SEP-2001.  
 PD  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.

XX (ELIT) ELITRA PHARM INC.  
 PA Hasebeck R, Ohlsen KU, Zyekind JW, Wall D, Trawick JD, Carr GU;  
 PI Yamamoto RT, Xu HH;  
 XX WPI: 2001-611495/70.  
 DR N-PSDB; AAS51470.  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX Example 3; Seq ID No 5107; 511pp; English.  
 PS  
 XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 4342 AA;  
 Query Match 3.1%; Score 121; DB 22; Length 4342;  
 Best Local Similarity 21.4%; Pred. No. 2.5;  
 Matches 178; Conservative 101; Mismatches 318; Indels 236; Gaps 44;  
 QY 4 YAVAA-----AVGVLLIAGAGN-----AGBARAAVRA-LVARKL 40  
 DB 1181 VAIQAERSPOLLVGLAIVKAGVYPLDPYSERLAYMLAAGVELLITQAHLERL 1239  
 QY 41 GPG-----PADEFSEVERALAAKPGDTSLSGGGGAARVVGSGVAAAAGLHRYLRD 95  
 DB 1240 -PGEAGVTPICDLSIKIDNWPQAGLHL--GDNLAYVYVTSGSTQPKGVG----- 1289  
 QY 96 FCGCHVAMSGSQLRPRPLPAVPGE--LTBATPNRYRYONVCTQSYSPVMDMARERE 153  
 DB 1290 --NTHAALA-ERLQWQATYTLDDGDVLMQKAP-----VSFVSVW--E 1328  
 QY 154 IDWMAALNGIMLALAMSGO-----EAIWQRYVLLGLTGAINEFFTGPAFLANGRWG 205  
 DB 1329 CFWPLVIGCCLVLAAPGEHDPALVZLVROFGTTTHFVPLPLQLFIDEEGVAAGSLR 1388  
 QY 206 NLHTWDPPLPPSWIKQLYLQHRVLDQMSFGMTPVLPAPAGVPEAVTRVFPQVNVYTK 265  
 DB 1389 RLFGSGEALPAE-----LRNRYLQR-----LPAVALHNRVGPET--AINVT-- 1428  
 QY 266 GSWHFNCSYSGSLAPED---PIPIIGSLFLRLITERGTDHLYGADTFNEM--- 317  
 DB 1429 ---HMOC-----RAEDGERSPIGRPLGNVVCRLDAEF---NLLPAGVAGELCTGG 1473  
 QY 318 -----OPSPSEPSYLAATTAAYVEAMTAVDTEAVMLLOGLM-----FCHQPOFWG- 362  
 DB 1474 IGLARVIGLRPAJAEKFAVDFPSAAGERLYRTGDRARNADAVLEVLGRLDQVLRGR 1533  
 QY 363 ---PAQIRAVLGAVP---RGRLLVLDLPASQPV--YTRTASQGGPFTWMLANFGGNH 414  
 DB 1534 RIEPEEIQARLLAQPVAQAVVIREGVAAGSLQVGYTGAVGAEA-----EQHQ 1584  
 QY 415 GLFALBAVNGGPEAAALFPNSTWVGTGMABEGISQNEVVYSIMAEIGWRK---DPVDD 470

DB 1585 RLRAALQA-----ELPEYVWPTQLMRLAQMPLGPS-GLIDTRALPEPWAQREHVEPTE 1638  
 QY 471 L-----AAWVTSFAARRYG-----VSHPDAGAMLLIRSVNCS-- 505  
 DB 1639 LQRRALAIWSEVLGIPRVGLDDDFELQGHSLIARIVSRTRQACDVLPRLALPEASEL 1698  
 QY 506 -----GEACRGHNRSPLVRPSPLOMNTSIWYNSRDVEAMRLLITSAPLSPAF 556  
 DB 1699 EAPCEQVBAQAAGRTDGHGAIIRIDRQPVPLSYQQRMWFLMQL-----EPD--SPAY 1751  
 QY 557 RY-----DLIDLR--QAVQELVSLYEARSAYLSKELASLRA---GSV-LAYELL 603  
 DB 1752 NVGILARISGLPUDVAFRPAALQALVQ-RHETLRTPSVVDGVQVRYHGDGLHMDQDF 1810  
 QY 604 PALD-----EVLASDSRFLGSLBQA---RAAAVSEADPYEQNSRYQLTWGPE 652  
 DB 1811 SALDSDSRQOHLQTLA-DSEARFPDLSSGPLLRVCVMKABREHYLVVTLHHIV---TE 1866  
 QY 653 GNILDYANKQLAGVANYTTRMRFLFALVDVAGGIPFGQHPDKVYQLE 705  
 DB 1867 GWAMDIFARELGALYEAPLDR-----ESPLEPLVQ--YIDYSVWQRE 1908  
 RESULT 6  
 ID AAM22601 standard; Protein; 4472 AA.  
 XX AAM22601;  
 AC AAM22601;  
 XX 27-FEB-1998 (first entry)  
 DT  
 XX  
 DE Tyactone synthase ORF1 protein.  
 XX Tyactone synthase gene cluster; tylG gene; multifunctional protein;  
 XX polyketide; tyactone synthesis; antibiotic; tyloisin.  
 KW Streptomycetes fradiae.  
 OS  
 XX  
 FT Key Location/Qualifiers  
 FT Misc-difference 1  
 FT /note= "encoded by GTG"  
 FT Domain 43..447  
 FT /note= "ketosynthase domain, KSq(s)"  
 FT Domain 586..914  
 FT /note= "acyltransferase domain, At(s)"  
 FT Domain 954..1038  
 FT /note= "acyl carrier protein domain, ACP(s)"  
 FT Domain 1060..1483  
 FT /note= "ketosynthase domain, KS1"  
 FT Domain 1606..1934  
 FT /note= "acyltransferase domain, Att1"  
 FT Domain 2199..2378  
 FT /note= "ketoreductase domain, KR1"  
 FT Domain 2469..2552  
 FT /note= "acyl carrier protein domain, ACP1"  
 FT Domain 2576..2999  
 FT /note= "ketosynthase domain, KS2"  
 FT Domain 3149..3477  
 FT /note= "acyltransferase domain, Att2"  
 FT Domain 3502..3687  
 FT /note= "denitratase domain, DN2"  
 FT Domain 4016..4200  
 FT /note= "ketoreductase domain, KR2"  
 FT Domain 4302..4385  
 FT /note= "acyl carrier protein domain, ACP2"  
 PN EP791655-A2.  
 XX 27-AUG-1997.  
 PD  
 XX 19-FEB-1997; 97EP-0301056.  
 XX





XX XX Homo sapiens.

XX XX Key Location/Qualifiers

XX XX Domain 575..827

XX XX Domain /note= "Eukaryotic protein kinase domain"

XX XX Domain 580..812

XX XX /label= Protein\_kinase\_domain

XX XX WO200196547-A2.

XX XX 20-DEC-2001.

XX XX 14-JUN-2001; 2001WO-US19444.

XX XX 15-JUN-2000; 2000US-212073P.

XX XX PR 23-JUN-2000; 2000US-213467P.

XX XX PR 30-JUN-2000; 2000US-215651P.

XX XX PR 07-JUL-2000; 2000US-216605P.

XX XX PR 13-JUL-2000; 2000US-218372P.

XX XX PR 25-AUG-2000; 2000US-228055P.

XX XX (INCY-) INCYTE GENOMICS INC.

XX PA Yye H, Lal P, Bardman O, Borowsky ML, Au-Young J, Lu Y;  
PI Gandhi AR, Trisouley CM, Wallia NK, Yao MG, Lu DAM, Greenwald SR;  
PI Rankumar J, Griffen JY, Kearney L, Burford N, Nguyen DB, Tang YT;  
PI Baughn MR, He A, Thornton M, Hafalia A, Patterson C, Gurrarajan R;  
PI Lo TP, Khan F, Reardon SA, Azimzai Y, Policky JL, Ding L,  
PI Grether M, Elliott VS, Thangavelu K, Batra S, Ison CH;

XX XX WPI; 2002-090207/12.

XX DR N-PSDB; AAD26467.

XX XX New polypeptides, useful for diagnosing, treating or preventing  
PT disorders of growth and development, cardiovascular and lipid, and  
FT diseases such as cancer, comprise human kinase polypeptides -  
PT

XX XX Claim 1; Page 164-165; 197P; English.

XX XX The invention relates to human kinase PKIN proteins and their  
CC corresponding cDNAs. A composition containing PKIN agonist is useful for  
CC treating a disease or condition associated with decreased expression of  
CC PKIN and a composition comprising PKIN antagonist is useful for treating  
CC a disease or condition associated with overexpression of PKIN. The  
CC disorders include cancer (leukaemia, adenocarcinoma, lymphoma, melanoma,  
CC myeloma, sarcoma, teratocarcinoma, Hodgkin's disease), immune disorder  
CC (Acquired Immune Deficiency Syndrome (AIDS), asthma, Addison's disease,  
CC attherosclerosis, anaemia, allergies, adult respiratory distress syndrome,  
CC autoimmune thyroiditis, gout, bronchitis, Crohn's disease, diabetes  
CC mellitus, multiple sclerosis, pancreatitis, psoriasis, Reiter's disease,  
CC osteoarthritis, osteoporosis, pancytopenia, uveitis, ulcerative colitis,  
CC rheumatoid arthritis, Sjogren's syndrome, warts, ulcerative infections)  
CC bacterial, parasitic, fungal, viral, protozoal and hematithic infections)  
CC Growth and development disorders (arteriosclerosis, cirrhosis, hepatitis,  
CC Cushing's syndrome, hypothyroidism, cerebral palsy, cataracts); cardio  
CC vascular disease (arteriovenous fistula, hypertension, vasculitis,  
CC aneurysms, congestive heart failure, angina pectoris, myocarditis,  
CC ischaemic heart disease, chronic bronchitis, lung tumours); lipid  
CC disorder (fatty liver, Fabry's disease, Niemann-Pick's disease,  
CC hypercholesterolemia, obesity). PKIN DNA is useful for assessing  
CC toxicity of a test compound and in gene therapy. The present sequence  
CC is human PKIN-20 protein.  
XX XX

SQ Sequence 871 AA;

Query Match 3.0%; Score 118.5; DB 23; Length 871;  
Best Local Similarity 18.6%; Pred. No. 0.34;  
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;



```

QY 19 GGAAGDEAREAAVALVARL-LGPGPAD---FSVSVERALLAAKGLD-----TY 65
DB 21 GSASQSSSSQVSLRVSSQVGTPEPSLDAEGMTQEAEDLSSTPTLQRPQVMTKRF 80
QY 66 SLGG-GGAARVVRGSGVAAAGLHRYLDFCG-----CHVANGSGQLR-----109
DB 81 SLGGRGYAGVAGYGTAFAGDAG-----GMLGGGPMARLMAVVSQSEEEQEARAE 134
QY 110 -----LP---RPLPAVPGELTEATPNRRYYQVNCQSYSFVMMDMARWE 151
DB 135 SQSEEQEARAEPLPOVSARPVDEVGAPTRSSP-----EPTPWE 175
QY 152 -----REI---DMALNGINLAL-----AMSGEAIWQRYIAL-----182
DB 176 DIGQVSLVQIRDLSDGAADDTISLDISEVDPAYLMLSDLYDKYLPFEFMTFRKVPKSA 235
QY 183 -----GLTQAEINER-----FTGP-----AFLAMGRMGNLHTWDGP 213
DB 236 QPEPSPMAEEELAEFPETPTWPGELGPHAGLETBESDVALLAEAAVGRKRMSSP 295
QY 214 LPPSWHIKQLYLQHRVLDQMRSPGMPVLPAPAGHV-----PEAVTRVPQVNVTKMG 266
DB 296 SRSIFHPPG---RHLPLDEPAELGLRRVKAAYEHISRLIKGRPEGKEKGP---RKTP 349
QY 267 SWGHFNCSYSCSFLAPEDPIFFIGSLFLRELKERGDHITGADTFNMQPPSSSPY 326
DB 350 GLASFRLSLGKXMDRAP-----TFUREL-----SDEITV 378
QY 327 LAAATTAAYEAMTAVIDEAVMLQGMFLQHQPOFWGPAQIRAVLGAVPFG---RLVLDLF 384
DB 379 LGQSVTLACQVSAQPAQATMSDQABLE-----SSSRVILISATLKNFOLLITLVVV 430
QY 385 AESGPVYTRTASFGQGFIMCMHNFGNNGLFGALEAVNGGEARLPPNGTMGTGA 444
DB 431 AEDIGVYI-----CSVSNALGVTYTTGVLR-----KAER--PSSS-----463
QY 445 PEGISQNEVVYSIMAEIWRKDPVPD-----LAAW---VTSFAARRYGVSHPDAGAA 493
DB 464 -----PCPDIGEYVADGVLLVWKPVSXGVTYTYOCSSLEGGS 501
QY 494 WRLLRSVYNC-----SGEAC 509
DB 502 WTLIASDIPFCYLTSLKLSRGTYTPTACVSKAGMPYSSPEEOVLLGGPSHAEBS 561
QY 510 RGHNRSLVVRPSLQWMTSIWYNR-SDVFEAMRLILTSAPSLATSPAFRYDLDLTR--Q 566
DB 562 QGRSAQPLBETKTAFTQIQGRFSYVRQWEKASGRALAAKIIPHPDKTAIVREYE 621
QY 567 AVQELVSLYVEARSAVLS-KEIASLLRAGGVLAPELLPALDEVLASDSRFLGSMLEQA 635
DB 622 ALKGLRHPHLAQHLAAVLSPRHLVILIEL--CSGPEILLPCIAE-RASYSESEVKDYLMQW 678
QY 626 RAAA-----VSEADRYEON---SRQYLTLMGEGEIIIDYANKQLAGVANYTTRMNL 677
DB 679 LSAIYVHNGHIHLDRSENMITEYNLL-----KVVDLGAQSLISQEKVLPDSDFKO 732
QY 678 FLEALVDSVAQG---IPFOQHOFKXVFCU-EOAFVLSKQRYPSQSPRG 721
DB 733 YLEMAEELLEGGQAVP-----QTDIMATGVAFIMLSAEVYVSSSG 774

```

```

KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
KW mitral valve disease; aortic valve disease; tricuspid valve disease;
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;
KW atherosclerosis; cardiac tumour; microbial infection.
OS Homo sapiens.
FN WO20063381-A1.
PD 26-OCT-2000.
PF 11-APR-2000; 2000MO-US09488.
PR 16-APR-1999; 99US-0129553.
PA (SCIO-) SCIOS INC.
PI Zeng W, Stanton L, Kong H;
DR WPI: 2001-007013/01.
DR N-PSDB; AAC62285.
PT Novel h1965 polypeptides capable of regulating signal transduction and
PT exhibiting kinase activity useful for identifying antibodies to treat
PT cardiac diseases, and additional mediators of signal transduction -
PS Claim 1; Page 55-57; 81pp; English.
FX
CC The present sequence represents a human protein with putative function
CC in signal transduction. The polypeptide is designated H1965. The protein
CC is capable of regulating signal transduction and exhibits kinase
CC activity. The H1965 transcript is expressed in the heart. H1965
CC polypeptides and polynucleotides are useful for preventing or treating a
CC cardiac disease, such as congestive heart failure, dilated congestive
CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,
CC mitral valve disease, aortic valve disease or tricuspid valve disease,
CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,
CC and cardiac tumours in humans. The polypeptide is also useful for
CC detecting the expression of a protein capable of regulating signal
CC transduction or the expression of a protein capable of acting as a donor
CC or acceptor molecule of a phosphate group. The monoclonal antibodies can
CC be used as probes for detecting discrete antigens expressed by tissue or
CC cell samples, and therefore used in humans for localization and
CC monitoring of microbial infection.
SQ Sequence 1351 AA;
Query Match 3.0%; Score 118.5; DB 22; Length 1351;
Best Local Similarity 18.6%; Pred. No. 0.69;
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;
QY 19 GGAAGDEAREAAVALVARL-LGPGPAD---FSVSVERALLAAKGLD-----TY 65
DB 501 GSASQSSSSQVSLRVSSQVGTPEPSLDAEGMTQEAEDLSSTPTLQRPQVMTKRF 560
QY 66 SLGG-GGAARVVRGSGVAAAGLHRYLDFCG-----CHVANGSGQLR-----109
DB 561 SLGGRGYAGVAGYGTAFAGDAG-----GMLGGGPMARLMAVVSQSEEEQEARAE 614
QY 110 -----LP---RPLPAVPGELTEATPNRRYYQVNCQSYSFVMMDMARWE 151
DB 615 SQSEEQEARAEPLPOVSARPVDEVGAPTRSSP-----EPTPWE 655
QY 152 -----REI---DMALNGINLAL-----AMSGEAIWQRYIAL-----182
DB 656 DIGQVSLVQIRDLSDGAADDTISLDISEVDPAYLMLSDLYDKYLPFEFMTFRKVPKSA 715
QY 183 -----GLTQAEINER-----FTGP-----AFLAMGRMGNLHTWDGP 213
DB 716 QPEPSPMAEEELAEFPETPTWPGELGPHAGLETBESDVALLAEAAVGRKRMSSP 775
QY 214 LPPSWHIKQLYLQHRVLDQMRSPGMPVLPAPAGHV-----PEAVTRVPQVNVTKMG 266

```

```

Db 776 SSSLHFPPG---RHLPDEPAELGLRERKASVEHISRLKCRPGLEKEGP---RKKP 829
QY 267 SMGHFNCSYSCSFLLAPEDPIFPIIGSLFLRELLEFGTDHIYADTNEMOPSSSEPSY 326
Db 830 GLASFRLSLGKSMWRAP-----TFRLREL-----SDETVV 858
QY 327 LAAATTAVTEAMTAVTEAVTWLLOGLFQHPQFQWGPQAQIRAVLGAVPRG--RLVLDLF 384
Db 859 LGQSVTLACQVSAQPAQAATWSKDGAPLE-----SSSRVLISATLKNFQLLTILVVV 910
QY 385 AESQPYVTRTASFOGQPFIFWCMILHNFQGNHGLFGALEAVNGGPEAARLFPNSTWGTGMA 444
Db 911 AEDLGVT-----CVSNALGVTYTTGVL-----KAEK--PSSS----- 943
QY 445 PEGISQNEVYVYSIMAEIGWRKDPVD-----LAAM--VTSFAARRGVSHPDAGAA 493
Db 944 -----PCPDIGEVADGVLLWPKPVESYGPVTYIVQCSLEGGS 981
QY 494 WRLLRSVYNC-----SGEAC 509
Db 982 WTTLASDITDCCITLSKLSRGTYTFRTRCVSKAGMGVSPSEQLGCSHLSAEESS 1041
QY 510 RGHNSPPLVRPSLQNNISIWYNR--SDVEAWRLILTSAPLSATSPAFRYLLDLTR--Q 566
Db 1042 QGRSAQPLPSTKTFAFQTQIQGRFSEVVRQCEKASGRALAKIIPYHKOKTAVLRBEYE 1101
QY 567 ANQELVSLYEERKSYLS--KELASLIRAGVLAWEILPALDEVIASRPLFSGWEIOA 625
Db 1102 ALKGLHPHQAQMAAYLSPRHLVILLEL--CSGPELPLCLAE--RATSYSESEVKDYLMQM 1158
QY 626 RAAA-----VSEAEADFEYGN--SRVQTLTWGPBGNILIDYANKOLAGLVANYTPRWRL 677
Db 1159 LSAITQYLHNOHILHLDLRSENMILITEYNLL-----KVVDLGNAGSLQSEKVLPSDKFKD 1212
QY 678 FLEALYDVSAAQ---IPFOGHPDKKVPOL--EQAFPLSKQKRPSPQPG 721
Db 1213 YLETPAPELLEGGAVP-----QTDIMAGVTAFTMLSAEYVPSSEG 1254

RESULT 11
AAB30568
ID AAB30568 standard; Protein; 1610 AA.
AC
XX AAB30568;
DT
XX 19-MAR-2001 (first entry)
DE
XX A full length human signal transduction polypeptide.
XX
XX Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
XX congestive heart failure; dilated congestive cardiomyopathy;
XX hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
XX mitral valve disease; aortic valve disease; tricuspid valve disease;
XX myocardial infarction; cardiac arrhythmia; arteriosclerosis;
XX atherosclerosis; cardiac tumour; microbial infection.
XX
OS Homo sapiens.
XX
XX WO200063381-A1.
XX
XX 26-OCT-2000.
XX
XX 11-APR-2000; 2000WO-US09488.
XX
XX 16-APR-1999; 99US-0129553.
XX
XX (SCIO-) SCIOS INC.
XX
XX Zeng W, Stanton L, Kong H;
XX
XX WPI; 2001-007013/01.
XX
XX N-PSDB; AAC62286.

```

```

XX
PT Novel h19G5 polypeptides capable of regulating signal transduction and
PT exhibiting kinase activity useful for identifying antibodies to treat
PT cardiac diseases, and additional mediators of signal transduction
PS Claim 1; Page 61-65; 81pp; English.
XX
CC The present sequence represents a human protein with putative function
CC in signal transduction. The polypeptide is designated H19G5. The protein
CC is capable of regulating signal transduction and exhibits kinase
CC activity. The H19G5 transcript is expressed in the heart. H19G5
CC polypeptides and polynucleotides are useful for preventing or treating a
CC cardiac disease, such as congestive heart failure, dilated congestive
CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,
CC mitral valve disease, aortic valve disease or tricuspid valve disease,
CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,
CC arterial or rheovascular hypertension, arteriosclerosis, atherosclerosis
CC and cardiac tumours in humans. The polypeptide is also useful for
CC detecting the expression of a protein capable of regulating signal
CC transduction or the expression of a protein capable of acting as a donor
CC or acceptor molecule of a phosphate group. The monoclonal antibodies can
CC be used as probes for detecting discrete antigens expressed by tissue or
CC cell samples, and therefore used in humans for localization and
CC monitoring of microbial infection.
XX
SQ Sequence 1610 AA;
Query Match 3.0%; Score 118.5; DB 22; Length 1610;
Best Local Similarity 18.6%; Pred. No. 0.9;
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;
QY 19 GGAAGDEBARBAAVRALVAPL--LQGPBPD--FSVSVERALAKGLD-----TY 65
Db 760 GSASQASSQVSSISRVGSSQVGTPEPSLDAEGMTQAEPLDSSTLTQRPQDQATWRKF 819
QY 66 SLIG--GGAARVRGSGVAAAGLHRYLDFCG-----CHVWMSGQLR----- 109
Db 820 SLGRRGGVAGVAGYTAFGDNG-----GMLCGQPMARIMANYSQSEEEQEBARAE 873
QY 110 -----LP--RPLPAVPEGLTEATPNRYRYQNVCTQSFSFWMDARWE 151
Db 874 SQSEEQEABAEPLPQVSARPVDEVRAPTRSSP-----EPTPWE 914
QY 152 -----REL--DMALNGINIAL-----AMSQGEALWQRYVAL 182
Db 915 DIGQVSLVQIRDSGDREAADTISLDISEVDPAVILNSDYDKILPFRSMIRKPKSA 974
QY 183 -----GLTQAEINEF-----FTGP-----AFLWGMGNILHTWDGP 213
Db 975 QPEPSPMAEBELAEFEPTWPMFGELOPHAGLEITESESDVDALLAEAVGRKRWSSP 1034
QY 214 LPPEWHIKQLYLQHRVLDQKRSQMTPLVLPAPAGV-----PEAVTRVFPQVNTYKMG 266
Db 1035 SRSLSHFPPG---RHLPDEPAELGLRERKASVEHISRLKCRPGLEKEGP---RKKP 1088
QY 267 SMGHFNCSYSCSFLLAPEDPIFPIIGSLFLRELLEFGTDHIYADTNEMOPSSSEPSY 326
Db 1089 GLASFRLSLGKSMWRAP-----TFRLREL-----SDETVV 1117
QY 327 LAAATTAVTEAMTAVTEAVTWLLOGLFQHPQFQWGPQAQIRAVLGAVPRG--RLVLDLF 384
Db 1118 LGQSVTLACQVSAQPAQAATWSKDGAPLE-----SSSRVLISATLKNFQLLTILVVV 1169
QY 385 AESQPYVTRTASFOGQPFIFWCMILHNFQGNHGLFGALEAVNGGPEAARLFPNSTWGTGMA 444
Db 1170 AEDLGVT-----CVSNALGVTYTTGVL-----KAEK--PSSS----- 1202
QY 445 PEGISQNEVYVYSIMAEIGWRKDPVD-----LAAM--VTSFAARRGVSHPDAGAA 493
Db 1203 -----PCPDIGEVADGVLLWPKPVESYGPVTYIVQCSLEGGS 1240
QY 494 WRLLRSVYNC-----SGEAC 509

```



QY	626	RAAA-----VSEADAFYQN---SRVQTLTWGPGENTLIDPANKQALGLVANYIYPRML	677
DB	1426	LSATQYLNHCHILHLDLRENNMILTEHYL-----KVYDGNAGQSUSQEKVLPSDKFD	147
OY	678	FLEALVDVAQG--IPFOQHGFKNVFL-EOAFVLISKQIPYSPQPRG	721
DB	1480	YLEETMAPELLGEGAVP-----QDIMAIGTAFIMLSAEYPPSSSG	1521
RESULT 13			
ID	AAE24151	standard; Protein; 1665 AA.	
XX	AAE24151;		
AC			
XX			
XX	23-SEP-2002	(first entry)	
DE		Human kinase (PKIN)-22 protein.	
XX			
KM	Human; kinase; PKIN; cancer; immune system disorder; atherosclerosis;		
KM	acquired immune deficiency syndrome; AIDS; Addison's disease; allergy;		
KM	aschem; multiple sclerosis; psoriasis; arteriosclerosis; cirrhosis;		
KM	development; hepatitis; cardiovascular; hypertension; drug screening;		
KM	myocardial infarction; Goodpasture's syndrome; lipid disorder; growth;		
KM	fatty liver; Gaucher's disease; Niemann-Pick's disease; anorectic;		
KM	hypercholesterolemia; obesity; gene therapy; cytostatic; anti-hiv;		
KM	neuroprotective; hepatotropic; hypotensive; cardiac; nephrotropic;		
KM	hyperlipidaemia; enzyme.		
XX			
OS	Homo sapiens.		
XX			
XX			
FF	Key	Location/Qualifiers	
FT	Domain	68..118	
FT		/note="Immunoglobulin domain"	
FT	Domain	165..418	
FT		/note="Eukaryotic protein kinase domain"	
FT	Domain	167..401	
FT		/note="Protein kinase domain"	
FT	Domain	1174..1235	
FT		/note="Immunoglobulin domain"	
FT	Domain	1369..1621	
FT		/note="Eukaryotic protein kinase domain"	
FT	Domain	1372..1606	
FT		/note="Protein kinase domain"	
FN	WO200233099-A2.		
XX			
PD	25-APR-2002.		
XX			
XX			
PF	20-OCT-2001; 2001WO-US47728.		
XX			
PR	20-OCT-2000; 2000US-247410P.		
PR	27-OCT-2000; 2000US-240688P.		
PR	03-NOV-2000; 2000US-245708P.		
PR	09-NOV-2000; 2000US-247672P.		
PR	16-NOV-2000; 2000US-249565P.		
PR	22-NOV-2000; 2000US-252730P.		
PR	01-DEC-2000; 2000US-250807P.		
XX			
PA	(INCYTE GENOMICS INC.		
P1	Gururajan R, Baughn MR, Walla NK, Elliott VS, Xu Y, Arvizu C;		
P1	Yao MG, Ramkumar J, Ding L, Tang YT, Hatala AJ, Nguyen DB;		
P1	Gandhi AR, Lu Y, Yue H, Burford N, Bandman O, Tribouley CM;		
P1	Lal PG, Recipon SA, Lu DAM, Borowsky ML, Thornton M, Swarnaker A;		
P1	Thangavelu K, Khan FA, Ison CH;		
DR	WPI: 2002-454603/48.		
DR	N-PSDB; AAD38655.		
XX			
PT	New human kinase polypeptide, for diagnosing, preventing and treating		
PT	cancer, immune system disorders, growth and development disorders,		
PT	cardiovascular disorders and lipid disorders		

XX Claim 1; Page 182-186; 210pp; English.

CC The invention relates human kinases (PKIN) and their corresponding  
CC nucleic acid sequences. PKIN and its DNA are useful for diagnosing,  
CC treating and preventing cancer, an immune system disorder (e.g.,  
CC acquired immunodeficiency syndrome (AIDS), Addison's disease, allergy,  
CC asthma, arteriosclerosis, multiple sclerosis, psoriasis), disorders  
CC affecting growth and development (e.g., atretosclerosis, cirtosis,  
CC hepatitis), cardiovascular disorder (e.g., hypertension, myocardial  
CC infarction, Goodpasture's syndrome), and a lipid disorder (e.g., fatty  
CC liver, Gaucher's disease, Niemann-pick's disease, hypercholesterolaemia,  
CC hyperlipidaemia, obesity), and for assessing the effects of exogenous  
CC compounds. Anti-PKIN antibody is useful in a diagnostic test for a  
CC condition or a disease associated with the expression of PKIN in a  
CC biological sample. A composition comprising PKIN or an agonist or  
CC antagonist of PKIN is useful for treating a disease or condition  
CC associated with decreased or increased expression of functional PKIN.  
CC PKIN is useful in a number of drug screening techniques and to analyse  
CC the proteome of a tissue or cell type. PKIN DNA is useful for creating  
CC knockin humanised animals or transgenic animals to model human diseases,  
CC and in somatic or germ-line gene therapy. The present sequence is human  
CC PKIN protein.

SQ Sequence 1665 AA;

Query Match 3.0%; Score 118.5; DB 23; Length 1665;  
Best Local Similarity 18.6%; Pred. No. 0.95;  
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

Dy 19 GGAGDAREARAVALVARL-LGPGAAD---FSVVERALAKPLD-----TY 65

Dd 815 GSASQAESSQSSTLRWSSQGTERPGLSAEGTGAEDLSPTLPLORPQEQTWRKF 874

Oy 66 SLGS-GGAARVRVGSGTVAAAGLHRYLDPCG----GVAVMSGSLR-----109

Dd 875 SLGGRGYAGVAGCTAFAGDGAG--GMIGQMARIANAVSSEBEDEBARAE 928

Oy 110 -----LP---RLPALVEGELTEATPRNRRYYQNVCYSYSTFWMIDNRWE 151

Dd 929 SASEGEQEARAESFLPVDSARPVEVERADPTRSP-----EPLPWE 969

Oy 152 ----REI-DMMANGIMIAL----AMSGOEAIVORYITAL-----182

Dd 970 DIGVSLVOIRDLSDGDMAADTLISDISVPAYLNSTDLYIKYLPEEFMIFRKVKESA 1029

Oy 183 -----GLTAQLNEF-----FGCP-----AFAMGMINHTMDGP 213

Dd 1030 QEPESPFAEBELAEPFEFTPMWFEGELGFPHAGEITESEDVDLLLEAAGRKRKKASP 1089

Oy 214 LPFMWHIKOLILQRVLWDOKRSFGMTFLVPAFAGHV-----PEATRFPQVVNTRMG 266

Dd 1090 SRLFFHPG--RHLPDPERAIGLRRRVASAVENTSRILMKRGPLEKEGRP---BKP 1143

Oy 267 SWGFNFNGSYSCSFLPADBDPIFLTISGLFLRELKERGTTHIYAADFNMENQPSSPSKY 326

Dd 1144 GLASFRLSGLKSWRAP-----TFIREL-----SDETLV 1172

Oy 327 LLAATTVAYTEAMTVDTAEVALLLGWLFOHQPOFWMPBAORIAVLGAVPRG--RLVTLDLF 384

Dd 1173 LGGSVTLACVVASAPAQATWSKGAPLE-----SSSHVISAILTKNQLTITLIYVV 1224

Oy 385 AESOPVTRKTASFCQQFIWCMNLIPGNHGHLFGALEVNNGPPAAPRLFPRSTWGICWA 444

Dd 1225 AEDUGVVI-----CSVANLAGITYTTGYLR-----KAER-BESS-----1257

Oy 445 PEGISIONEVVYSLMALGWKRDVPVD-----LAAM-VTSFAARRCVSHPDAGA 493

Dd 1258 ---PCPDIGEYADVAGVILLVMKVESYGPTYTIYCSSLREGS 1295

Oy 494 WRLLLRSVNC-----SGEAC 509

Dd 1296 WTLIASDIFFDCCYLTSKISRGIYTTPRTACYCKXMGKPYSSPQOVYLLGSSHIASEBS 1355

```

OY 510 RGNRSPLVRPSLTQNTSIWYNR-SDVFEARLLLTASPLATSPAFRYDLDTL-TR--Q 566
DB 1356 QGRSAQPLPSTKTPAFQTOIGRFSVVRQCEKASGRLLAKIIPHPKDTAVLREYE 1415
OY 567 AVQELVSLTYEERASAYLS-KELASILPAGVLAAYELLPAIDVLAUSDSPFLGSMLEQA 625
DB 1416 ALKGLRHPHLAQTLAAAYLSPRHVLILEL--CSGPELLPCLAE-RASYSESEVXDYLWQM 1472
OY 626 RAAA-----VSEAEADFYEQN---SRVQTLTMGPENILIDYANKQLAGLVANYTTPRRL 677
DB 1473 LSAIQYLHNOHILHDLRSENNITETNYL-----KVVDIGNAQSLSQEKVLPDKEFD 1526
OY 678 FLEALVDSVAQG---IPFOHQFDKNVFOI-EQAFVLSKORVPSQPRG 721
DB 1527 YLETAPELLEGGAVP-----QTDIWAIGVTAFLMSAEVPSSEG 1568

RESULT 14
AAOI5372
ID AAOI5372 standard; Protein; 1665 AA.
AC AAOI5372;
DT 19-SEP-2002 (first entry)
XX 19-SEP-2002 (first entry)
DE Human myosin light chain kinase subfamily-related kinase protein.
KW Human; gene therapy; chromosome 1; kinase protein;
KW myosin light chain kinase subfamily; kinase protein-mediated disease;
KW transgenic animal.
OS Homo sapiens.
XX WO200240683-A2.
XX 23-MAY-2002.
XX 22-OCT-2001; 2001WO-US32616.
XX 14-NOV-2000; 2000US-0711134.
XX 17-MAY-2001; 2001US-0858664.
XX (PEKE ) PE CORP NY.
XX Wei M, Ketchum K, Di Francesco V, Beasley EM;
XX WPI; 2002-500223/53.
XX N-PSDB; AAL43908, AAL43909.
XX New kinase proteins related to myosin light chain kinase subfamily and
XX encoding polynucleotide, useful for diagnosing, treating disease or
XX condition mediated by the kinase protein and for identifying modulators
XX
XX Claim 1; Fig 2; 96BP; English.
XX The invention comprises the amino acid and coding sequences (located on
XX chromosome 1) of a human kinase protein that is related to the myosin
XX light chain kinase subfamily. The human kinase DNA and protein sequences
XX of the invention are useful for identifying agents that modulate the
XX activity of the human kinase protein. Kinase-modulating agents are useful
XX for treating a disease or condition mediated by a human kinase protein.
XX The human kinase DNA sequences can be used to produce transgenic animals
XX which are useful for studying the function of kinase proteins and
XX identifying/evaluating modulators of kinase protein activity. The present
XX amino acid sequence represents the human kinase protein of the invention.
XX
XX Sequence 1665 AA;
XX
XX Query Match 3.0%; Score 118.5; DB 23; Length 1665;
XX Best Local Similarity 18.6%; Pred. No. 0.95;
XX Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

```

```

OY 19 GGAAGDEAREAAVAVLVARL-LGEPDAD-----FSVSEVALAAKPGD-----TY 65
DB 815 GSAASQASSQVSLVAGSQVETEPGLDAGMTQGEHEDLSDFPTLQRPQVYTKRF 874
OY 66 SLGG-GGAARVVRGSTGVAAGLHRYLRDFCG-----CHVAGSGQLR-----109
DB 875 SLGRRGGVAGVAGYGTFFAGDAG-----GMLGGQPMWARIAMAVSOSSEEEQEARAE 928
OY 110 -----P-----RPLPAVPGELTEATNRRVYQNVCTQSYSEVWMDMAWE 151
DB 929 SOSSEQGEARASPLQVSARVPEVGRAPTRSS-----EPTWE 969
OY 152 -----REI--DMALNGLNAL-----AMSGQALWQRYVAL-----182
DB 970 DIGQVSLVQIRDLSDAEADITSLDISVDPAVNLSDLYDKLPFEFMIFRVPKSA 1029
OY 183 -----GLTQALINF-----FTGP-----AFLANGRMGNLHTWGP 213
DB 1030 QEPSPMAHEELAEFPETFWPBGELGHALEITESESDVDALLAEAVGRKRWSSP 1089
OY 214 LPPSWHIKQVLYQHRVLDQMRSGMTPVLPAPAGHV-----PEAVTRVPQVAVTKWG 266
DB 1090 SRLTFHPG---RLPLDEPAELGLRERYKASVEHISRLKGRPEGLEKEGPP---RKXP 1143
OY 267 SMGHFNCSYSCSFLAPEDPITIGSLFLRLIKEFGTDHLYGADTFREMQPPSSPEBY 326
DB 1144 GLASFRLSGLSKMDRAP-----TFLRL-----SDEVTV 1172
OY 327 LAAATTAVEAMTAVDTEAVWLLQGMFLFOHQPOFMPQAPRAVLGAVPRG--RLVLDF 384
DB 1173 LGQSVTLAQVSAQAQAATWKGDAPE-----SSRVLISATLKNFGLTILVYV 1224
OY 385 AASQVYRTTASFOQOPFIWCLHNFQGNHGLFGLLEAVNGSPEARLIPNSTMTGTGMA 444
DB 1225 AEDLGYTT-----CSVSNAIGVTTVTGLR-----KAER--PSSS-----1257
OY 445 PEGISQNEVVYSLMELGWRKDPVPD-----LAAM--VTSPAARVYGHSDPAGAA 493
DB 1258 -----PCPDIGEVADSVLWTKPVESTGYPTIYVQCSLEGG 1295
OY 494 WRLLRSVYNC-----SGEXC 509
DB 1296 WTLTASDIPDCCYLTSKLSRGQTYTFRUACVSKAGGFPSSPEQVLLGSPHLSSEBS 1355
OY 510 RGNRSPLVRPSLTQNTSIWYNR-SDVFEARLLLTASPLATSPAFRYDLDTL-TR--Q 566
DB 1356 QGRSAQPLPSTKTPAFQTOIGRFSVVRQCEKASGRLLAKIIPHPKDTAVLREYE 1415
OY 567 AVQELVSLTYEERASAYLS-KELASILPAGVLAAYELLPAIDVLAUSDSPFLGSMLEQA 625
DB 1416 ALKGLRHPHLAQTLAAAYLSPRHVLILEL--CSGPELLPCLAE-RASYSESEVXDYLWQM 1472
OY 626 RAAA-----VSEAEADFYEQN---SRVQTLTMGPENILIDYANKQLAGLVANYTTPRRL 677
DB 1473 LSAIQYLHNOHILHDLRSENNITETNYL-----KVVDIGNAQSLSQEKVLPDKEFD 1526
OY 678 FLEALVDSVAQG---IPFOHQFDKNVFOI-EQAFVLSKORVPSQPRG 721
DB 1527 YLETAPELLEGGAVP-----QTDIWAIGVTAFLMSAEVPSSEG 1568

RESULT 15
AAB30569
ID AAB30569 standard; Protein; 2596 AA.
XX AAB30569;
XX
XX 19-MAR-2001 (first entry)
XX
XX A splice variant of a signal transduction polypeptide.
XX Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;

```

KM congestive heart failure; dilated congestive cardiomyopathy;  
 KM hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
 KM mitral valve disease; aortic valve disease; tricuspid valve disease;  
 KM myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
 KM atherosclerosis; cardiac tumour; microbial infection; splice variant.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT Domain 325..504  
 FT /note="guanine nucleotide exchange factor domain"  
 FT Domain 1094..1351  
 FT /note="kinase domain"  
 FT Domain 2301..2553  
 FT /note="kinase domain"  
 XX  
 PN W020006381-A1.  
 PD 26-OCT-2000.  
 XX  
 PF 11-APR-2000; 2000MO-US09488.  
 XX  
 PR 16-APR-1999; 99US-0129553.  
 XX  
 PA (SCIO-) SCIOS INC.  
 XX  
 PI Zeng W, Stanton L, Kong H;  
 XX  
 DR WPI; 2001-007013/01.  
 DR N-PSDB; AAC62287.  
 XX  
 PT Novel h19G5 polypeptides capable of regulating signal transduction and  
 PT exhibiting kinase activity useful for identifying antibodies to treat  
 PT cardiac diseases, and additional mediators of signal transduction -  
 XX  
 PS Claim 1; Page 68-74; 81pp; English.  
 XX  
 CC The present sequence represents a splice variant of human in signal  
 CC transduction polypeptide. The polypeptide is designated H19G5. The  
 CC protein is capable of regulating signal transduction and exhibits kinase  
 CC activity. The H19G5 transcript is expressed in the heart. H19G5  
 CC polypeptides and polynucleotides are useful for preventing or treating a  
 CC cardiac disease, such as congestive heart failure, dilated congestive  
 CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,  
 CC mitral valve disease, aortic valve disease or tricuspid valve disease,  
 CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,  
 CC arterial or rheovascular hypertension, arteriosclerosis, atherosclerosis  
 CC and cardiac tumours in humans. The polypeptide is also useful for  
 CC detecting the expression of a protein capable of regulating signal  
 CC transduction or the expression of a protein capable of acting as a donor  
 CC or acceptor molecule of a phosphate group. The monoclonal antibodies can  
 CC be used as probes for detecting discrete antigens expressed by tissue or  
 CC cell samples, and therefore used in humans for localization and  
 CC monitoring of microbial infection.  
 CC  
 SQ Sequence 2596 AA;  
 XX  
 Query Match 3.0%; Score 118.5; DB 22; Length 2596;  
 Best Local Similarity 18.6%; Pred. No. 1.9; Indels 319; Gaps 40;  
 Matches 165; Conservative 97; Mismatches 307;  
 QY 19 GGAAGDEARBAAVRALVRL-LGPGPAD--FSVSEVERLAKPGLD-----TY 65  
 DB 1746 GSASQASSSVSSIRVSSQVGTPEPSILDAGMTQEAEDLSDSTPTLQRPQOATMRKF 1805  
 QY 66 SLGG-AGAAVRYVSGTAAAGLHRYLRDPG--CHVAMSGSLR-----109  
 DB 1806 SLGGRGYAGVAGTAFAGDAG-----GMLGGPMARIKAVSOSSEBEQEARAF 1859  
 QY 110 -----LP-----RPLPAVPGELTEATPNRRYRYONVCTQSYSPVWMDMARWE 151  
 DB 1860 SQSEEQGEARAEAPLPQVARSARVPEVGRAPTRSSP-----EPTPWE 1900

QY 152 -----REI--DWMALNGINLAL-----AMSGQZAIWQRYVAL-----162  
 DB 1901 DIQVSLVQIRDLSDGDEADDTSLDISEVDPAVNLNLDYIKYLPEFEMI PRKPKSA 1960  
 QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGMLHMTWDP 213  
 DB 1961 QPEPSPMAEBELAEFEPEPTWMPGELGPHAGLEITEESBDVALLAEAAVGRKKKSSP 2020  
 QY 214 LPPSMHIKQLYLQHRVLDQKRSFGMPFVLPAPAGHV-----PEAVTRVPQVNTXMG 266  
 DB 2021 SRSLFHPPG---RHLPLDEPAELGLRRERKASVEHISRLIKGRPEGEKGP---RKKP 2074  
 QY 267 SWGHFNCSYSCGFLLAEDPIPIIGSLFRELKERGTHTHYGADTFNMQSPSSPSY 326  
 DB 2075 GLASFRLSGKSWDRAP-----TFPLREL-----SBETVV 2103  
 QY 327 LAAATTAVEAMTAVDTEAVWMLQGMWLFQHPQFQWGPAGIYRAVAGVPSG--RLVLDLF 384  
 DB 2104 LGQSVTLACQVSAQPAQAATMSKDGALE-----SSSRVLISATLKNFOLLTLVV 2155  
 QY 385 AESQPVYTRTASFGQGFPIWCMLEHFGNHLFGALBAVNGGPEAPARLFNSTMTVGTMA 444  
 DB 2156 AEDLGVT-----CSVSNALGVTITGVLR-----KAER--PSSS-----2188  
 QY 445 PEGISQNEVYVSIAMELGWKKDPVPD-----LAAM--VTSFARRYGVGHPDAGAA 493  
 DB 2189 -----POPDIGEYVADGVLLVMKPVESYGPVYTYVQCSLEGS 2226  
 QY 494 WRLLRSVYNG-----SGEAC 509  
 DB 2227 WTLASDIFDCCVYLSKLSRGTYFRTACVSKAGMGPSPSEQVLLGAPSHLASEES 2286  
 QY 510 RGHRSPLVRRPSIQMNTSIWNR--SDVFEAMRLLTSASLATSAPFRYDLDTLP--Q 566  
 DB 2287 QGRSAQPLPSTKTAFTQTQIRGRFSVVRQCKEASGRALAAKTIIPHPADOKTAVLREYE 2346  
 QY 567 AVCELVSLYVEEARSAVLS--KEIASLRAGGVLAAYELLPALDEVLASDSRFLGSMWEQA 625  
 DB 2347 ALKGLRHPHLAQHAAVLSRHLVLIEL--CSGPELLPCLAE--RASYSSEVVDYLMQM 2403  
 QY 626 RAAH-----VSEADADYEON--SRQYLMPGEGNILDYANKQLAGVANYTTPWRLL 677  
 DB 2404 LSAQYLNQHILHLDRSENMITEYNLL-----KVVDLGNAGSISQEKVLPSSDKFKD 2457  
 QY 678 FLEALVDSVAGQ---IPFOQHOFKXVFOU--EQAFVLSKORYSQPFG 721  
 DB 2458 YLEIMARPELLBGQAVP-----QTDIMALGVTAIFMLSAEYPVSSBG 2499

Search completed: February 13, 2004, 16:15:23  
 Job time : 60 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: February 13, 2004, 15:44:39 ; Search time 29 Seconds

(without alignments)  
2463.908 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939

Sequence: 1 MEAVAVAAVAVVLLAGAG.....VDLAKKIFLKRYPGWAGSW 743

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3939	100.0	743	1 G02270	alpha-N-acetylgluc
2	1060.5	26.9	770	2 G87316	alpha-N-acetylgluc
3	123.5	3.1	497	1 G02938	probable debrisoqu
4	121	3.0	4342	2 H83343	probable non-ribos
5	119	3.0	3649	1 S18268	delta-(L-alpha-aml
6	117.5	3.0	497	1 Q4HDD1	debrisoquine 4-hyd
7	117	3.0	1374	2 D72593	hypothetical prote
8	115.5	2.9	973	2 T35238	probable secreted
9	115	2.9	725	2 C87485	ComEC/Rec2 family
10	114	2.9	479	2 T31293	aldelyde dehydroge
11	113	2.9	1281	2 G83405	hypothetical prote
12	112	2.8	1323	2 I78557	N-methyl-D-asparta
13	111	2.8	1027	2 H87316	hypothetical prote
14	109.5	2.8	285	2 F87285	conserved hypochet
15	109.5	2.8	814	2 T47641	hypothetical prote
16	109.5	2.8	1377	2 G55159	thsa protein precu
17	109.5	2.8	1411	2 B55145	thsb protein precu
18	109	2.7	1080	2 T19048	probable pro-x car
19	107.5	2.7	1622	2 D86428	glutathione S-conj
20	107	2.7	741	2 A83271	hypothetical prote
21	106.5	2.7	666	2 A87577	oligopeptid trans
22	106.5	2.7	1678	2 T35547	hypothetical prote
23	104.5	2.7	726	1 CSECHP	catalase (EC 1.11.
24	104.5	2.7	726	2 G81237	hydroperoxidase HP
25	104	2.6	1230	2 T07663	soluble starch syn
26	103.5	2.6	726	2 C86085	catalase, hydroper
27	103.5	2.6	1737	2 T00209	MEGF8 protein - hu
28	103	2.6	698	2 D90771	hypothetical prote
29	103	2.6	698	2 H85633	hypothetical prote

30	103	2.6	1400	2 B70963	hypothetical prote
31	102.5	2.6	1623	2 T01369	ABC transporter At
32	102	2.6	352	2 B87507	conserved hypochet
33	102	2.6	435	2 T35884	tagatose 6-phospha
34	102	2.6	9376	2 T14593	syringomycin synth
35	101.5	2.6	638	2 AD3111	myo-inositol catad
36	101.5	2.6	638	2 H98175	myo-inositol catad
37	101.5	2.6	843	2 F96537	hypothetical prote
38	101.5	2.6	3194	2 D71517	toxoin-like outer m
39	101.5	2.6	5069	2 T17464	rifamycin polyketi
40	101	2.6	698	2 F64839	yea protein precu
41	101	2.6	709	2 B82580	alanyl dipeptidyl
42	101	2.6	3413	2 T17467	rifamycin polyketi
43	101	2.6	3519	2 S43048	polyketide synthas
44	100.5	2.6	4735	2 T17463	rifamycin polyketi
45	100	2.5	726	2 A10936	catalase (hydroper

## ALIGNMENTS

## RESULT 1

G02270  
alpha-N-acetylglucosaminidase (EC 3.2.1.50) - human  
C/Species: Homo sapiens (man)  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #ext\_change 03-Dec-1999  
C/Accession: G02270  
R/Weber, B./ Scott, H./ Hopwood, J.J.  
submitted to the EMBL Data Library, November 1995  
A/Reference number: H00952  
A/Accession: G02270  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-743 <WEB>  
A/Cross-references: EMBL:U40846; NID:g1197839; PID:g1197840  
A/Genes: GDB:NAGLU  
C/Genetics:  
A/Cross-references: GDB:636533  
A/Map position: 17c21  
A/Note: deficient in Sanfilippo B syndrome  
C/Superfamily: human alpha-N-acetylglucosaminidase  
C/Keywords: glycosidase; hydrolase

Query Match	100.0%	Score 3939	DB 1	Length 743
Best Local Similarity	100.0%	Pred. No. 5e-277		
Matches 743	Conservative	0	Mismatches 0	Indels 0
			Gaps 0	
QY	1 MEAVAVAAVAVVLLAGAGAGDPAARBAANRALVARILGPPAADTSVSVERLAAPK 60			
DB	1 MEAVAVAAVAVVLLAGAGAGDPAARBAANRALVARILGPPAADTSVSVERLAAPK 60			
QY	61 GLDTYSLGGGAARVRVSGSTGVAAAAGHRLRFCCGCHVAVMSGQRLPRPLAVPGE 120			
DB	61 GLDTYSLGGGAARVRVSGSTGVAAAAGHRLRFCCGCHVAVMSGQRLPRPLAVPGE 120			
QY	121 LREATPNRRYYQNYCTOSYSFVWMDMARWEIEDMALLNGINALAMSGDAIQRVYL 180			
DB	121 LREATPNRRYYQNYCTOSYSFVWMDMARWEIEDMALLNGINALAMSGDAIQRVYL 180			
QY	181 ALGLTQAEINERFPTAPFLAMGRMGLHTWQDPLPSSHIIKQLYCHVLDQMSFGMT 240			
DB	181 ALGLTQAEINERFPTAPFLAMGRMGLHTWQDPLPSSHIIKQLYCHVLDQMSFGMT 240			
QY	241 VLPAPAGHVEAVTRVFPQVNTKXSGHFNCSYCSFLAPEDPIPIIGSLFRL 300			
DB	241 VLPAPAGHVEAVTRVFPQVNTKXSGHFNCSYCSFLAPEDPIPIIGSLFRL 300			
QY	301 KEFGTDHYTGADTFEMQPPSSSEPSYLAATTAAYEANTAVDTEVWLQGLPQHQOF 360			
DB	301 KEFGTDHYTGADTFEMQPPSSSEPSYLAATTAAYEANTAVDTEVWLQGLPQHQOF 360			
QY	361 WBPQIRAVLGAVPGRLLVLDLFAESQPVYTRTASFOGPIFWMLHFGNHGIFGAL 420			
DB	361 WBPQIRAVLGAVPGRLLVLDLFAESQPVYTRTASFOGPIFWMLHFGNHGIFGAL 420			



```

Db      361  MGAQIRAVLGAVERGRLLVLDLFAESQPYTRTASFGQGFIMWMLHNGHGLFGAL 420
QY      421  EAVNGGEARLFPNSTMTGTGMAPEGISONEVYISLMAELGRKDPVVDLAAMTSPFA 480
Db      421  EAVNGGEARLFPNSTMTGTGMAPEGISONEVYISLMAELGRKDPVVDLAAMTSPFA 480
QY      481  RRVGSHPDGAAMRLLRGVNCSGEGAGHNRSLYVRPSLQMTSITWYNSDYFEAM 540
Db      481  RRVGSHPDGAAMRLLRGVNCSGEGAGHNRSLYVRPSLQMTSITWYNSDYFEAM 540
QY      541  RLILTSAPSLATSPARFYDLIDLTROAVQELVSLYEARSAYLSKELASLLRAGCVLAY 600
Db      541  RLILTSAPSLATSPARFYDLIDLTROAVQELVSLYEARSAYLSKELASLLRAGCVLAY 600
QY      601  ELIPALDEVLASDRFLSGSWLEQARAAVSEADPFYEQNSRYQULTMGPEGNILDYAN 660
Db      601  ELIPALDEVLASDRFLSGSWLEQARAAVSEADPFYEQNSRYQULTMGPEGNILDYAN 660
QY      661  KOLAGLVANYTTPRMRLFLBALVDVSAQGIFFQOHQFDKNVFLQEQAFYLSKQRYPSQPR 720
Db      661  KOLAGLVANYTTPRMRLFLBALVDVSAQGIFFQOHQFDKNVFLQEQAFYLSKQRYPSQPR 720
QY      721  GDTVDLAKKIFLYKYPGMAVGSW 743
Db      721  GDTVDLAKKIFLYKYPGMAVGSW 743

```

## RESULT 2

```

C87316
Alpha-N-acetylglucosaminidase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: C87316
B:Nieman, W.C.; Fealduyem, T.V.; Paulsen, I.T.; Nelson, K.B.; Eisen, J.; Heidelberg, J.
R.; Laub, M.T.; Debey, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Kojot
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87316
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-770 <STO>
A:Cross-references: GB:AE005673; NID:g13421729; PIDN:AAK2527.1; GSPDB:GN00148
C:Gene: C00540
C:Superfamily: human alpha-N-acetylglucosaminidase

```

```

Query Match      26.9%; Score 1060.5; DB 2; Length 770;
Best Local Similarity 33.0%; Pred. No. 1,2e-68;
Matches 260; Conservative 133; Mismatches 302; Indels 93; Gaps 19;

```

```

QY      2  EAVAVAAVGVLLIAGAGAGDEBRRAAVALVARLLGPGPADFSVSVERBALAAKGG 61
Db      11  QALSLAAAVASPALAAGSTDG---VAAARASLKRKLFGRRLAG-----AHLVTTPG 59
QY      62  LDT--YSLGGGGAARVVRSGTGVAAAAGLHRYLRDFCGGVHVMSSGQLRLEPRLPAVPG 119
Db      60  AENSWVIGKGGK--IISGDSPALVRGAYAHLRQAGLAHVWEDRVVQAGAVPAGG 118
QY      120  ELTEAPFNRYRYQNVCTGSYFVWMDWAEREIDIMMALINGINLALANSQGEALWQRY 179
Db      119  ARVE--TFPRHAYLNTCTGYTTPTWGMGWRTEIWMMAHGIDMLAMEGGEYVFRALM 177
QY      180  LAIQLTQAEINEFTGPAFLAMGMGNLHTWDGPLEPSPWHIKQLYLQHVILDMQSFQMT 239
Db      178  REFSLSAELADYFSGAFTPMHMGNIIGYKAPLPAIDKXKDLQVKILGMSLGMT 237
QY      240  PVLPAFAGHYPEAVTRVFPQVNTYKMGSWGFNCYSQSFLLAPEDPIPIIGSLFLREL 299
Db      238  PILPAFGYVPAFAEKNPARYRYRMPEGFHETY---WLDPADPLFAKIAARLALY 293
QY      300  IKRGTHIYGADITFNMQPP-----SSEPSY-----LA 328

```

```

Db      294  TETGAGTYIADSFNEMLEPINADGADARDAVQSTANTAVYTKKVEVDPALKAQRLA 353
QY      329  AATTAVENTAVDTEAVMLIQMLFQHQPFQMGPAQIRAVLGAVERGRLLVLDLFAESQ 388
Db      354  AYGAIVDSIRKOTRPAVVMWQGLFQADSHFMDPAISAYSLVLPDDKLMILIDIGDRY 413
QY      389  P-VYTRTASFGQGFIMWMLHNGHGLFGAL-----EAVNGGEARLFPNSTMTVG 440
Db      414  PNWKAKAFEGKRWITGYHANNYSNPVYGDGFYQDIPALANFDACKL-----AG 467
QY      441  TGMAPESIQNEVYISLMAELGRKDPVVDLAAMTSPFAARVGVSHPDGAAMRLLR 500
Db      468  FGMPEGLHNSIYVEAVYDLAMSEQASP-ATWLTYYAARVYGKTSFALDALGQLVFA 526
QY      501  VYNGSGEACGHNSP-----LYVRPSLQMTSITWY--NRSDFEAMLLTSPSLATSP 554
Db      527  AFSTRVMSPRWKSQAQAYLFKKPIATVGDFQHPGDRAKLEAAVYALTLALPTYGQPR 586
QY      555  AFRYDLIDLTRQAVQELVSLYEARSAYLSKELASLLRAGVLAELLPALDEVLASDS 614
Db      587  LFVLDLTDATFRLATMKIDILLQVAVAAYRRGDTA---AGDAARVSI-----EALALSI 637
QY      615  RFLG-----SWLEQARAAVSEADPFYEQNSRYQULTMGPEGNILDYANKOLAGLV 667
Db      638  DKLLGVQDPTLATVIDEARFAGDTPADAAIVAAKAKQVITWGGSGMLNDYASAKWGLY 697
QY      668  ANYTTPRMRLFLBALVDVSAQGIFFQOHQFDKNVFLQEQAFYLSKQRY-----PSQPRGDT 723
Db      698  KSFLPFRMSRLDLDL--KAAGTGFDEVYTRGVAMERAMVEAVYREKPADPICEI 755
QY      724  VDLAKKI 730
Db      756  KTLIARI 762

```

## RESULT 3

```

G02938
Probable debrisoquine 4-hydroxylase (EC 1.14.14.-) cytochrome P450 - crab-eating macaqu
C:Species: Macaca fascicularis (crab-eating macaque)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C:Accession: G02938
R:Lawton, M.P.; Laddison, K.J.; Speirs, A.A.; Mankowski, D.C.; Tweedie, D.J.
submitted to the EMBL Data Library, October 1995
A:Reference number: G12616
A:Accession: G02938
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-497 <LAW>
A:Cross-references: EMBL:U08218; NID:g1022899; PIDN:AAA79722.1; PID:g1022900
C:Gene: CYP2D7
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: Chromoprotein; heme; iron; metalloprotein; monooxygenase; oxidoreductase; t
F/302-465/Domain: cytochrome P450 homology <P45>
F/443/Binding site: heme iron (Cys) (axial ligand) #status predicted

```

```

Query Match      3.1%; Score 123.5; DB 1; Length 497;
Best Local Similarity 22.9%; Pred. No. 0.45;
Matches 104; Conservative 47; Mismatches 140; Indels 163; Gaps 22;

```

```

QY      343  TEAVVLL-----QGLFQHQPFQMGPAQIRAVLGAVERGRLLVLDLFAESQPYTRT 394
Db      12  TVALFLLVLDLMHRROHMAARYPP---GPLFLPGL-----GNLHVD----- 50
QY      395  ASFGQPRIMWMLHNGFN-----HGLFGALBA-VNGGEARL--FENST 437
Db      51  --FKNTYFCQLRRRGVNSQLANTPPVYVNLGAAREALVTCEDPTADRPVYINQ 108
QY      438  MVGTGMAPEGI-----SQNEVYISLMAELGRKDPVVDLAAMTSPFAARVGVSH 487
Db      109  VLGQPRSQGVFLARYGPAWREGQRFVSSTLRNIGLK---KSLQWVTEBAACLCAAFT 165

```

QY 488 PDAGAMR--LLRSVNCSEACRGNHNSPLVRPSLQNTS:TWYNSDVFEAMRLLL 544  
 DB 166 DOAGRPFRPNSLDDAVSN-----VYASLYGR----- 193  
 QY 545 TSAPLANSPPARYD-----LLDLTRQAVOELVSLYEARSAYLSLAS-----LLR-- 593  
 DB 194 -----REYDDPRLFLFDLTHLEKE-----BSGFRLVLAIPLLKIP 234  
 QY 594 --AGVLAYE--LLPALDEVL-----ASDRSFLGSLWLEQARAAVAESADPYEQ 640  
 DB 235 GLAGVLSQKAFLLQDLBELTEHRTWDPQAPRPDLTEAFLEMEKAK--GNPSSFNEE 293  
 QY 641 NSRYQLTMGPENGLIDPAKQAGLVANYTTPRWRLEALVDSVACGIFPOOHQDPKN 700  
 DB 294 NLR-----MVVADLFSAQMTSTTLAAGLLMLTHPDVQRRV--QOEIDDV 338  
 QY 701 VFQEQAFVLSKORYP-----SQPRGDTVDL 726  
 DB 339 IGQVRPBMQDQARMPYTTAVTHVQRGDIYPL 372

## RESULT 4

H83343  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: H83343  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; Llm,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; PMID:20437337; PMID:10984043  
 A:Accession: H83343  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-4342 <STO>  
 A:Cross-references: GB:AE004663; GB:AE004091; NID:g9948460; PIDN:AA05812.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA2424  
 C:Superfamily: acyl carrier protein homology; acetate-CoA ligase homology  
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
 F:60-553/Domain: acetate-CoA ligase homology <AC11>  
 F:584-652/Domain: acyl carrier protein homology <AC12>  
 F:1174-1622/Domain: acetate-CoA ligase homology <AC12>  
 F:1637-1705/Domain: acyl carrier protein homology <AC12>  
 F:2232-2689/Domain: acetate-CoA ligase homology <AC13>  
 F:2708-2773/Domain: acyl carrier protein homology <AC13>  
 F:3759-4230/Domain: acetate-CoA ligase homology <AC14>  
 F:4248-4316/Domain: acyl carrier protein homology <AC14>  
 F:1669,2738,4280/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 3.1%; Score 121; DB 2; Length 4342;  
 Beef Local Similarity 21.4%; Pred. No. 14;  
 Matches 178; Conservative 101; Mismatches 318; Indels 236; Gaps 44;  
 4 VAVAA-----AVGTLTAGAGG-----AGBARAAVRA-LVARLL 40  
 DB 1181 VAICERSPQLLVGLALIVKAGAVPLDPYPSRLAIVMLADSGVELLLQALFRL- 1239  
 QY 41 GPG-----PAADFVSVERALAAKRGDLTYSLGGAARVVRSGTGYAAAAGIHRILRD 95  
 DB 1240 -FGAGVTPLCLDSIKLNNWPSQAFGLH--GNLAVITTSQSTQPKQVG----- 1289  
 QY 96 FCGCHVANSQQLRPLRPAPVGE--LTEATPNRYRYQVNTQSYFVWMDWARERE 153  
 DB 1290 --NTTALA-ERLQWQATYTLIDGVDVMQAK-----VSFVSVW--E 1328  
 QY 154 IDPMALANGINIALAWSG-----EAIQWRVIALGLTQAEINFEFTGPAFLARWG 205  
 DB 1329 CFWPLVTCRLVLAPEGHRDPARLVELVROFGVTTLHFVPLQLFIDEGVAACSLR 1388

QY 206 NHTWQDPLPSPWHIKQLYQHRVLDQMSFGMTPLYPAFAGVPAVTRVFPQVNTKX 265  
 DB 1389 RLFSGEALPAE-----LNRVLQR-----LPVALHNRYPTEI--AINVT-- 1428  
 QY 266 GSWGHNCSSCSFLAPD-----PIPIIGSLFLBELKEPOTDITGADTFNEN----- 317  
 DB 1429 ---HMOC-----RAEDGERSPIGRPLGVNVCVLAERF--NLLPAGVAGELCIG 1473  
 QY 318 -----QPSSEPSVLAATTAVYEAMTAVTEAVMLQGL-----FOHOPWG- 362  
 DB 1474 LGIARVGLRPAISARFVADPESAAGERLYRTGDARANNAGVLETLRGDQVLRGF 1533  
 QY 363 ---PAQIRAVLAVP---RGLIVLDLPAESQPY--YTTASFOGQPIWCMHNFQNA 414  
 DB 1534 RIEPEIQARLLAQPVAAVAVVIRGVAGSQGVYTTAVAGAEAA-----EQHQ 1584  
 QY 415 GLFGALAVNGGPEARLPFNSTMTVGTGAPEGISONEVYYSIMAEIGMK-----DPVPD 470  
 DB 1585 RLPAALQA-----ELDEVVPTQLMRLAQMPGPS--GKLDITALLPFPVQOQREHFRTE 1638  
 QY 471 L-----AAVTSFAARPYG-----VSPDGAAMRLLSVNC-- 505  
 DB 1639 LQRIIAIWSVVLGPRVGLRDPFELGHSILATRIVSRTQACDVELPALFEASGL 1698  
 QY 506 -----GEACRGNHNSPLVRPSLQNTS:TWYNSDVFEAMRLLLSAPSATSAPF 556  
 DB 1699 EAFCEQVRAQAAGRDSHGAIIRIDREPVPVLSQGMFLWL-----EPD--SPAY 1751  
 QY 557 KY-----DLIDLTR--QAVQELVSLYEARSAYLSKELASLLRA---GGV-LAYELL 603  
 DB 1752 NVGILARLSPDLVAFEAALQVQ--RHETRTTPPSVDGVQVHGDDGILHMDWQDF 1810  
 QY 604 PLND-----EVLASDRFLLGSWLEQA---RAAASADADFEQSRQVLTIMGPE 652  
 DB 1811 SALDRDSRQOHLQTLA-DEEAPRPDLGEGPLIRVCVMAREHRLVTLHHIV--TE 1866  
 QY 653 GNILDYANKQAGLVANYTTPRWRLEALVDSVACGIFPOOHQDPKNVFOLE 705  
 DB 1867 GMMMDIFAREIGALYEAFLDDR-----ESPLEPLVQ--YLDYSWQRE 1908

## RESULT 5

S18268  
 delta-(L-alpha-aminoadipyl)-L-cysteiny-L-valine synthetase - Streptomyces lactamdurans  
 C:Species: Streptomyces lactamdurans  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Nov-2000  
 C:Accession: S18268; B38171  
 R:Martin, J.F.  
 submitted to the EMBL Data Library, January 1991  
 A:Reference number: S18268  
 A:Accession: S18268  
 A:Molecule type: DNA  
 A:Residues: 1-3649 <MAR>  
 A:Cross-references: EMBL:X57310; NID:g45005; PIDN:CAA40561.1; PID:g45006  
 R:Coque, J.U.R.; Martin, J.F.; Calzada, J.G.; Liras, P.  
 Mol. Microbiol. 5, 1125-1133, 1991  
 A:Title: The cephamycin biosynthetic genes pcbAB, encoding a large multidomain peptide  
 genes in Acetomycetes chrysogenum and Penicillium chrysogenum.  
 A:Reference number: S15283; PMID:92065808; PMID:1956290  
 A:Accession: S15283  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 224-940;1319-2010;2373-3307 <COQ>  
 A:Cross-references: EMBL:X57310  
 A:Note: the source is designated as Nocardia lactamdurans  
 R:Coque, J.U.R.; Liras, P.; Laiz, L.; Martin, J.F.  
 J. Bacteriol. 173, 6258-6264, 1991  
 A:Title: A gene encoding lysine 6-aminotransferase, which forms the beta-lactam precurs  
 A:Reference number: A38171; PMID:92011390; PMID:1917857  
 A:Accession: B38171  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-23 <CO2>

A:Cross-references: GB:S57006  
 C:Genetics:  
 A:Gene: pchAB  
 C:Superfamily: alpha-aminoacyl-tRNA synthetase; acetyl-CoA ligase homolog  
 C:Keyword: carrier protein; cephamycin biosynthesis; phosphopantetheine; phosphoprotein  
 F:786-758/Domain: acetate-CoA ligase homology <ACLI>  
 F:786-856/Domain: acyl carrier protein homology <ACPI>  
 F:1392-1844/Domain: acetyl carrier protein homology <ACIP>  
 F:1862-1932/Domain: acetyl carrier protein homology <ACIP2>  
 F:2446-2895/Domain: acetate-CoA ligase homology <ACLI3>  
 F:2942-2980/Domain: acyl carrier protein homology <ACIP3>  
 F:520,1896,2944/Binding site: phosphopantetheine (Set) (covalent) #status predicted

Query Match 3.0%; Score 119; DB 1; Length 3649;

Best Local Similarity 21.1%; Pred. No. 16;

Matches 174; Conservative 63; Mismatches 283; Indels 306; Gaps 37;

16 AGGGAAGDARAAAVRALV-----ATLGGPAAADSVSEVALAKPGLDYSIGG 69

1988 AALPGALGLVRRHPLRLTKTDGVRROYFIPADVRLV----- 2031

70 GGAARVRGSGTVAAGLHRYLDFCCGVASGSQLRPPLPAYGSLTEATPNRY 129

2032 -----STVDRALDELVT-----RAGYVRLHELP-----RAEFPHGD 2070

130 RYQNVCTGYSFVWMDARMERET-----DMALGICINIALMSGOEAIWORY 179

2071 EYLSVAVHSCPDGSMDFRRELNALLDGVPEADLGALRGT-----YGFVAVRQY 2124

180 LALGLQAEINEFPTAPFLAMGRMGLT-----MDGFLPSPMHK-----OLYGHRYLD 231

2125 LT-GKLLALTEWT-----GALGFETALPDHPFPFDRGRNELSELDERITE 2176

232 QMSFGMT----- 239

2177 ALRELARTARVSLVYLGMCLMNMVYGOHDTGVTSPARGRPEEPRAVGFANLLA 2236

240 -----PVLPAFAGHVEAVTRVPOVNVTKMSGMGHFNCSYCSFLAEDPI-FPI 290

2237 LKRVDPAAITLPAVYASVEAAV-YAAQV-----GELPREOLVKELKEKDSRHP 2287

291 TGLPLRELKEKGDHIGADTFENMOPSSSEPSYLAATTAAYEAMTAVDTEAVWLQ 350

2288 LQNFLLQNV-----SDH-----TSALTGYQDS-----GGMTTKFD-LSATWTERATGLA 2333

351 GML-----FOHQPFMPQAQIRAVLAVGRGLVLDLFAESQ-FVY 391

2334 GNLTYAASFDDTSASGFIATFQHV-----LAEFASAAQITPILQITLIDEPQALPDA 2388

392 TRTASFQGFPTWCMLHFGNGHGLFGALEAVNGCFEARLFPNSTWVGTG---MAPEG 448

2389 TRRARPPGP-----GRCTRLE-----EVAATMPDVAVAVHGVRLTYREL 2430

449 SQ-NEVYSLMAELMKRDP-----VPLAAMTSPARRRYGVSHPDAGAA 493

2431 NESANLHRLHRSVAPRADELIALVLDKSELTVAIILAVWGAAYPIDPSYDDRIA 2490

494 W-----RLLRSVNCGEACRGNRGPVLRBPLQNTSITWNRSPVFAMLLTTS 546

2491 FMLSDFGALV-----AGEAHGSRVRC-----LTSGVLDLBDQDITG 2529

547 APSLATSPAFRYDLDTROAVGELVSLYEEARSAVLSKEIASLIRAGVLAAYELLPL 606

2530 EP-AANP-----VTETSTETELAAVAY-----TSQTKPKPAVIVSHOSV----- 2567

607 DEVLADSPFLGSMLEQARAAVSEADFEQNSRYOLTMGEGNTL----- 656

2568 DSRRAQSGRYFSPPESAIVFL---ANYVPDEVEGLALSLVGHGLVLPSSAAD 2624

657 ---DYANKQLAGLVANYTTPRWRLFLALVDSVAGIIPFOHQPD 698

2625 PAVELANRE-GL-----SYLSGTPTQVERED 2650

# RESULT 6

CAHND1  
 debrisoquine 4-hydroxylase (EC 1.14.14.1) cytochrome P450 2D6 - human

N:Alternate names: CYP2D6; cytochrome P450 isozyme 2D; cytochrome P450db1

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence, revision 31-Mar-1992 #text change 03-Mar-2000

C:Accession: S01199; A28883; J04156; A33629; A30335

R:Gonzalez, F.J.; Skoda, R.C.; Kimura, S.; Umeno, M.; Zanger, U.M.; Nebert, D.W.; Gelbo

Nature 331, 442-446, 1988

A:Title: Characterization of the common genetic defect in humans deficient in debrisoqu

A:Reference number: S01199; PMID:8122614; PMID:3123997

A:Accession: S01199

A:Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-497 <CON>

A:Cross-references: EMBL:X08006; NID:G30450; PIDD:CAA30807.1; PIDD:G30451

A:Title: Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino acid seque

A:Reference number: A28883; PMID:88314109; PMID:3410476

A:Accession: A28883

A:Molecule type: mRNA

A:Residues: 1-497 <CON2>

A:Cross-references: EMBL:M20403; NID:G181349; PIDD:AAA52153.1; PIDD:G181350

R:Jiang, Q.; Voigt, J.M.; Colby, H.D.

Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995

A:Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16):

A:Reference number: J04153; PMID:95251703; PMID:7733969

A:Accession: J04156

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-497 <KIM>

R:Kimura, S.; Umeno, M.; Skoda, R.C.; Meyer, U.A.; Gonzalez, F.J.

Am. J. Hum. Genet. 45, 889-904, 1989

A:Title: The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identificatio

A:Reference number: A33629; PMID:90072069; PMID:2574001

A:Accession: A33629

A:Molecule type: DNA

A:Residues: 1-373, 'V', 375-497 <KIM>

A:Cross-references: EMBL:M33388; NID:G181303; PIDD:AAA53500.1; PIDD:G181304

R:Manns, M.P.; Johnson, E.F.; Griffith, R.J.; Tan, E.M.; Sullivan, K.F.

J. Clin. Invest. 83, 1066-1072, 1989

A:Title: Major antigen of liver fibrosis: microsome autoantibodies in idiopathic autoimmu

A:Reference number: A30335; PMID:89155788; PMID:2466049

A:Accession: A30335

A:Molecule type: mRNA

A:Residues: 125-373, 'V', 375-485, 'T', 487-497 <MAN>

A:Cross-references: EMBL:M24499; NID:G522194; PIDD:AAA6403.1; PIDD:G522195

C:Genetics:

A:Gene: GDB:CYP2D6

A:Cross-references: GDB:132127; OMIM:124030

A:Map position: 22q13.1-22q13.1

A:Antons: 60/3, 118/1, 169/1, 222/3, 281/3, 329/1, 391/3, 439/1

C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology

C:Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metall

F:302-465/Domain: cytochrome P450 homology <CYP>

F:443/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 3.0%; Score 117.5; DB 1; Length 497;

Best Local Similarity 21.7%; Pred. No. 1.2; Indels 153; Gaps 22;

Matches 97; Conservative 53; Mismatches 144;

345 AVML-----QGWFQHPQFQWPAQIRAVLGVPRGRLVLDLFAESQFVYRTAS 396

14 AIFLLVLDLWERRQRWAAVYP-----GLPLPGL-----GNLLHVD----- 50

397 FQGPFTWCLHNFGR-----HGLFGLER-VNGGEBAARL---FPNSTWY 439

51 FQNTPYCFDQLRRRFQGVFSLQIATPVVVVNLGLAVRELVLHGEDTAPRPVPTIQL 110

440 GTGWAPRPGI-----SQNEVYVSLMAELMKRDPVDLAAMWTSFAARRYGVSHPD 489

```

Db      111 GFGRSQCVTLARGPANREORRPSVSTLRNLGCK--KSLQEWTEBA-----158
Qy      490 AGAAMRLRLRSVYNGSGEACRGNHNSPLVRRPSLQMTSTIWNNSDYFEAMRLTLTSAPS 549
Db      159 -----CLCAAFANHSGRPF--RPNGLDKAV-----SNVI-----AS 188
Qy      550 LATSPAFRYD-----LLDLTRQAVQELVSLIYEANSAVLSKSLASLIRAGVLAIE--601
Db      189 LTCGRFFRYDDPRFLRLDLAQEGKESGFLREVLNAVPLIHPAL--AGKYLROKA 246
Qy      602 LLPALDEVL-----ASDSRFLGSMLEQARAAVAEADPFYQNSRYQLTLWGPE 652
Db      247 FLTQDLDELTEHRMTWDPAQPPRLTEAFLEMERKAK-GNPSSFPNDENLR-----296
Qy      653 GNILDYANKQLAGLVANYTYTPRMRLFLBALVDSVAQGIPIQOQFDKNVFOLEQ-----706
Db      297 ---IVADLPFAGMVTSTTLAMGLIMLTHPDQRRV---QQRIDDIVIGVRRPEMGDQ 350
Qy      707 -----AFVLSKORYSPQPRGDTVDL 726
Db      351 AHMYTTAVIHEVQRF-----GDIVPL 372

```

## RESULT 7

D72593  
hypotheetical protein APE1213 - Aeropyrum pernix (strain K1)

```

C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: D72593
R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamitsu, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; KDNA Res. 6, 83-101, 1999
A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
A/Reference number: A72450; MUID:99310339; PMID:10382966
A/Accession: D72593
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1374 <KAM>
A/Cross-references: DDBJ:AP000061; NID:gs104821; PIDN:BAAB0202.1; PID:dl043968; PID:gs10
A/Experimental source: strain K1
A/Genetics:
A/Gene: APE1213

```

Query Match 3.0%; Score 117; DB 2; Length 1374;  
Best Local Similarity 20.0%; Pred. No. 5.6;  
Matches 163; Conservative 78; Mismatches 249; Indels 324; Gaps 37;

```

Qy      1 MEAVAVAAAVGVLLIAGG-----AAGDEAREAAVRA-----LVAR 38
Db      620 IDSTLETGVPHFIAGNGGPGGLGTVPATARLAVVAATDMAVLSLQPGYLPILAG 679
Qy      39 LLGPADPFSV-SVERLAAKPGIDTYSIGGGGAARVAVRGSTGVAAAGLHRYLDFC 97
Db      680 LGGYCDPXFBSARGSHGAKPKGL-----AATGFRYTTGRS 717
Qy      98 GCHVAMSGSQLRPLPLPAVPGELTEATPNRYRYQNVCTQSYSFVWDMAREREIDMW 157
Db      718 LDH--YTGGRLD-PRAAALFLFGTSMATP-----M 744
Qy      158 ALNLTNLTAL-----AMSGGEAI-----MCHVYLAIGLQAEINEFTGPAFLAMGMLHT 209
Db      745 AAGAAALTIQALKESLVERLGLBEWMLVYALSNLTAQO-----WRG---LPMGMNA---793
Qy      210 WDGLPSPMHK-----QVYLOHRYLDQMR-----SPGM-----TPV 241
Db      794 --GIYDAGAIRLTLGVQGLVYSATILBBAQAGVAPGYGIPALIVAGSGVETPV 851
Qy      242 LPARAGHPAATVTRFPQVNTTKMSWGHNCSSCSLILAPDIPPIIIGSLFLRELIX 301
Db      852 DIVLEGSPVALKAAEPLLEVTVSS-----VEILSSPVG 887
Qy      302 EFGTDHIGADTFNEMCPSSSEPSYLAATAVVAEAMTAVDTEAVWLQGLWFOHQOPW 361

```

```

Db      888 EGGYTMNNAASIDBSMLPOGP---VENSILTPYE-----VPDRG---REERTSW 932
Qy      362 GPAQIRAVL-----GAVPRGRLLVLDLFAESQPVYTRTASFQOQPIWMLHNF 410
Db      933 WEGYVYGVALLIYMDLDDGGLGDGEFYILIDIKSSNVF-----RVE-----976
Qy      411 GGNHGLFGLLEAVNGGPEARLFPNSTWYGTMAPBGISQNEVYSLMAELGW---RKDP 467
Db      977 -----LSNQEALAGREA-----LQMS-EGVAREEVLLRLALGLSMAGGTTBA 1019
Qy      468 VPDLA--WYTFEAPARRVGVSHPDAGAAWRLILRSVYNGSGEAC-----509
Db      1020 TVEIAAISVESBA-----LPLPAIVEGSAVVAEAVNPPEPGYITGF 1063
Qy      510 -----RGNRSPLV---RRPSL-----QNTSTIWNRSDFV 537
Db      1064 IVASTGVGEYRMPYTLISPYRSHGVHVPQGLGPEYEEAMVGAFFYSWMYEDGD--1121
Qy      538 EAMRLTLTAPSLATSPAFRYDLDLTRQAVQELVSLYEARSAYLSKSLIRAGGV 597
Db      1122 --WRIVPELPGGGLAVARLHPVEDGREAATNIDAY-----LYKNPSITVSEAGEV 1173
Qy      598 L--AVELLPALDEVILASDSRF--LLGSYLE-----CARAAVEAEADPFYQNSRYOL 646
Db      1174 VVGVEEMALASREASAPARRFPDPTLGFMDQSGSGEIVLAAVNA-----1220
Qy      647 TLMGPEGNILDYANKQLAG-----LVANYTTP 673
Db      1221 ---PGRVLLVYRSVQYSGETARQPVITLYHTP 1250

```

## RESULT 8

T35238  
Probable secreted cellulase - Streptomyces coelicolor

```

C/Species: Streptomyces coelicolor
C/Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C/Accession: T35238
R/Seeger, K.D.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A/Reference number: Z21572
A/Accession: T35238
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-973 <SEE>
A/Cross-references: EMBL:AL031515; PIDN:CAA20643.1; GSPDB:GN00070; SCOEDB:SC5C7.31C
A/Experimental source: strain A3(2)
A/Genetics:
A/Gene: SCOEDB:SC5C7.31C

```

Query Match 2.9%; Score 115.5; DB 2; Length 973;  
Best Local Similarity 18.4%; Pred. No. 4.4;  
Matches 165; Conservative 93; Mismatches 293; Indels 347; Gaps 45;

```

Qy      1 MEAVAVAAAVGVLLIAG-----AGGAAGDEAREAAVRLVRLIGPPADPFSVVERA 55
Db      225 MGSASDSTVGVTAAGPTVASPQQLGVQGESGTYEV---KLSQGTAVAVTTSBA 280
Qy      56 LAAPGLDTSYSGGG-----DARVAV-----RSGTGVAAAGLHRYLDFCG 98
Db      281 -SGNTGL--TLTGASLFTPSNMDTAQRVYVSADAGSGSAVVESTAPGHG-----329
Qy      99 CHVAMSGQLRLPRPLPA-----VPGELTEA-----TPNRYRY-----QNV 136
Db      330 -KAAVTVTOLAARVDAFLELYGKITDPANGYSPGIPYHSVETLIVEAPDGHETT 388
Qy      137 TOSYFVWMDMAREREIDMNLNGINILALAMSGGEAIWQRYVALGLQAEINEFTGP 196
Db      389 SEAYSTILM-----LQANVGRK-----TGD 408
Qy      197 AFLAMGRGNLHTWDGELPSPWHIKQVYQHRVLDQMSFGVTPVLPAPAGVPAVAT-R 255
Db      409 ---WSKFNN-----AMEIMETVMIPTHADQPTNSSYNASKP--ATYAPSLDTFN 452

```

```

QY 256 VEPQVNVTKMGSMGHFNCSYCSFLIAPEDPIFFIIGSLFLRELKEFTDHIYGA----- 311
DB 453 RFP-----APIDGTVSAGSDPIAGELMSAIGTIDVYGMHLQ 489
QY 312 ---DTFN-----EMOPSSSEPSYTL-----AAATTAVENTAVIDTER-----VW 347
DB 490 DVDNFTGYGNSPKCKAGPSDGTGSPSYINTFORAGQESVWEIVPOPTCDAFKYGKNGYLD 549
QY 348 LLOG-----WLFQHPQFQWGPACIIRAVLAGVPRGRLLVLDLFAESQ-----PVYRTT 394
DB 550 LFTGDSYAKQKFTAPD-----ADARAYQAA-----YKADIMAGQCKSEISATIDK 599
QY 395 ASFGQGPFTWCMHNFNGHGLFGALEAVNGEBAARLFPNSTWGTGMAPEGISQNEVY 454
DB 600 AAKCGDYLRAMEDKYFKYKG-----NCGVPSACPGATGKDSH 638
QY 455 YSLMABLMGRKQDVPDLAAVMTSPARARYGSHPDAGAWRLLLSYVNCSEACRGNR 514
DB 639 Y-----LISWYIWMG-----GAVDTSGAMWRI-----GSHHGGYQ 671
QY 515 SPLVR-----RPSLOMNTSIWYNSD-VFEAWRLLLTS-----APSLAT 552
DB 672 NPLAAVALSTDADLTKKSAATGQSDMAKSLDRQVEFYRMLQDEGALAGATNSMAGRAT 731
QY 553 SPAPFRDLDTQAVQELVSLYEE-----ARSAVLSKELASILLRAGVLA 599
DB 732 FPA-----GTFPYGMYYDEKPYHDPSPNQMGFOQMSNERVAEYQSG-- 777
QY 600 YEILPALDEVLASDSRFLIGSWLEQARAAVSEADPFYEQNSRYOLTV-----W- 649
DB 778 -----DAGAAKAVLDKWDVMALESTVNPDTF-----RIPSTLQMSGQPDYMA 821
QY 650 ---GPRGN---LLDYANKQLAGIVA-----NYTPR-----KALFEALVDV----- 686
DB 822 SSPGAGDLDHEVADYATND--VGVAAYAKXTLTYVADRSGDTEASTAALLDGMENQ 879
QY 687 -AAGIPFOHQFQKVFQLEQAFVLQSKQRYPSQPRADTVDLAK-----KIFLKYYPGW 738
DB 880 DALGIAVPEIRADYRNF--DDGIYVSGMSGIMPNQDITVDASTFASISFYQDDPAW 935

RESULT 9
ComC/Rec2 family protein [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: C87485
R/Nerman, W.C.; Feildlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: C87485
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-725 <STO>
A/Cross-references: GB:AE005673; NID:g1342335; PIDN:AKX3879.1; GSPDB:GN00148
C/Genetics:
A/Name: CCI904

Query Match 2.9%; Score 115; DB 2; Length 725;
Best Local Similarity 20.6%; Pred. No. 3.2;
Matches 134; Conservative 67; Mismatches 234; Indels 214; Gaps 29;

```

```

QY 101 -VAMSGQNL-----PRPLAVGELTEATNRYQNVCTQSYF---VWMDARWE 151
DB 170 PVKPGQATGLRAMLGPPPPAPG-----AYDFADAMFD----- 205
QY 152 REIDMAMLNINLALAMSGEALWQRYVYLAGLTOAINEFFGPAFLMGRMGNHTWD 211
DB 206 -----SIGGVGPAIGDIGEVTLDQPP-LRLRLVNA-VNAFRMDLQRLARGPSG 257
QY 212 GPLPSPWHIKQLYOHRVLDQMSFGMTPLPAFAGHVEAVTRVEPQVNVTKMGSMGHF 271
DB 258 GAAMVGH--EAMISEQNTAMRASGLAHILSIGHM--AIVGQVFW-VVRMGV----- 308
QY 272 NCYSYCSFLIAPEDPIFFIIGSLFLRELKEFTDHIYGDITNEMOPSSSESYLAAT 331
DB 309 -AAMPMLALAPGKILAASG-----LVSLGYLIVSG-----PPAEBAITASV 354
QY 332 TAV-----YEANTAVDTAIVMLQGM-----LFG 355
DB 355 AFLAILFDRRAITLHGALALALISILLKPEPAGEPGQMSFAVTAALVALAESMPKPVRE 414
QY 356 HQPQFW--GPAQIRAVLAGVPRGRLLVLDLFAESQPYTTRTASFQGP-----IWCMLN 409
DB 415 LSTFWMTQGPQALATWL-AVISAASLVAGL-----ATAFPMQHENVAVW----- 459
QY 410 FGNHGLFGALEAVNGEBAARLFPNSTWY-----GTGMAPEGISQNEVYSLMAELG 462
DB 460 -----GLPAILAVSPSSFYIMEFLAIGYLFEEFGLSA-----PFLAVAG 499
QY 463 WRKQVPDLAAVMTSPARARYGVSHPDAG-----AKRLIASV----- 501
DB 500 WGIQVLMNVA-----GLPSEAHGQHTIVASAPQVLVAFLGLMVLCLMRGRMIGAPLA 555
QY 502 -----YNCSGEACRGHNSPVRSPSLQMTNTSIWYR 533
DB 556 LAVALMFRPFPDPAWIMADGITAIVRSQNAVALRTDPAKFGAELMARR 604

RESULT 10
aldehyde dehydrogenase homolog - Sphingomonas aromaticivorans plasmid pNL1
11293
C/Species: Sphingomonas aromaticivorans
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C/Accession: T31293
R/Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Senses, C.W.;
submitted to the EMBL Data Library, July 1998
A/Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromac
A/Reference number: Z20992
A/Accession: T31293
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-479 <ROM>
A/Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378434; PIDN:AAD04017.1
C/Genetics:
A/Genome: plasmid pNL1
A/Note: nmr

Query Match 2.9%; Score 114; DB 2; Length 479;
Best Local Similarity 23.3%; Pred. No. 2.1;
Matches 112; Conservative 52; Mismatches 165; Indels 152; Gaps 24;

```

```

QY 386 ESQPYTTRTASFOGPFITWCLHNFNGHGLFGALEAVNGGEAARL.FPNSTWVTGMA 444
    |||
    |||
Db 202 ---YV-----GNGAEIG--DVLIGHKPKVASITFTGSSRGVKAIA 235
QY 445 PEGISONEVYSLMELGWRKDPVPLDAMWTSFAARRGVSHPDAGAMRLLSRYNC 504
    |||
    |||
Db 236 -EIAARNLKITYL--ELG-GKSPVYCA-----DADLDKAVNAALFSTIMY 277
QY 505 GSEACGHNRSPLVRRPRLQWNTSIWYNSDVEEAMRLLTSPSLATSFAFRYDLDDLT 564
    |||
    |||
Db 278 QGVCMGASR-----IYERS-IFDQF---TFAFAATGRANSGDLRDP 318
QY 565 -----RQVQELVSLYEBARSAYLSKELASLRAGVLAVELLPALDEVLASDSRFL 617
    |||
    |||
Db 319 TMLGPIISERDRVRPHRIDARS-----KGAAVLAG-----350
QY 618 LGSWLEQARAAYSE---AEADFEQNSRYQLTWGPEGNI---LDYANKQLAGLVANY 671
    |||
    |||
Db 351 -GEMSGNSCAATILSGVTAMTVPEEFTGPTSLFPFDLTLEALNNNEYGLSASIF 409
QY 672 T 672
    |||
Db 410 T 410

```

## RESULT 11

```

G83405
Hypothetical protein PA1923 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83405
R:Stover, C.K.; Pham, X.O.; Ewitt, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Vann, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; Lim,
., Loty, S.; Olson, M.V.
Nucleotide 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: G83405
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1281 <STO>
A:Cross-references: GB:AE004618; GB:AE004091; NID:g9947912; PIDN:AA005311.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1923

```

```

Query Match 2.9%; Score 113; DB 2; Length 1281;
Best Local Similarity 22.5%; Pred. No. 9.8; Mismatches 141; Indels 174; Gaps 26;
Matches 104; Conservative 44;
QY 349 LOGWLFQHQPFQWGPQAIRAVLAGAVP-----RGRLLVLDLFAESQPYVTRTA--SFQG 399
    |||
    |||
Db 40 LAGMARAEAGVELRG--LRIGIGEARPGEMLDGNTLLIDTRPDRQAVEALGERIQG 96
QY 400 --OPTFWCLHNFNGHGLFGALEA-----VNGGP-----EAARLPFNSTWVG 440
    |||
    |||
Db 97 GTQPMI-----RVGGPPPGFGNLPALGRLVGYVYANGEANLRRLFEAVRRMAGLTVD 151
QY 441 TGMAPGIGISO-----NEVYSLMAE--LGMKCDVPDLAAVNT-----SFAARVYGS 487
    |||
    |||
Db 152 ALPAPQPLAQAGFYHPDAPPAVLAADYLAW-----GASRWASAPRLAFLIRGAIAD 205
QY 488 PDGAAMRLLRSVYNSGEGACGHNRSPLVRRPRLQWNTSIWYNSDVEEAMR-----541
    |||
    |||
Db 206 AQTGAIDELIRRS-----ERHQAFL-----AVFPDSDP--EALRKSFGA 245
QY 542 --LLITSASLATSFAFRYDLDDLTROAVQELVSLYEE-----ARSAVLSKELASL 591
    |||
    |||
Db 246 DVQALVNLQHLQNGPARAEFLALD--VPVLTQIGYRDGNADQWLAASGVAPRTAAAF 302
QY 592 -----LRAGVLAVELLPALDEVL-----ASDSRELLGSW 621
    |||
    |||
Db 303 LGMPETWMSDPLVITSALENGEPKLMAG--QAEALDDKDLRLRLRRLPADKHLALMFW 360

```

```

QY 622 L-----EQAPAAVSEAEADFEQNSRYQLTWGPEGNIIDY 658
    |||
    |||
Db 361 NHPEGENVAASHLNPASLARLGEALRAAGYVAVNS-----ESALIDT 405
QY 659 ANKQLAGLVANYTTPRRKLFLEALV-DVAQGIPEQO--HQFD 698
    |||
    |||
Db 406 AQRLLGQ-----YYRPQ--TLDALYRDGLAASLPLDAYLHWF 441

```

## RESULT 12

```

178557
N-methyl-D-aspartate receptor chain NMDAR2D-2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 21-Jan-2000
C:Accession: I78557; I58158; D45219
R:Monyer, H.; Burnashev, N.; Laurie, D.J.; Sakmann, B.; Seeburg, P.H.
Neuron 12, 529-540, 1994
A:Title: Developmental and regional expression in the rat brain and functional properti
A:Reference number: I58158; MUID:94206533; PMID:7512349
A:Accession: I78557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1323 <RES>
A:Cross-references: GB:I31612; NID:g469068; PIDN:AAC37647.1; PID:g469069
A:Accession: I58158
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-66, 'V', 68-1323 <RE2>
A:Cross-references: GB:I31611; NID:g469066; PIDN:AAC37646.1; PID:g469067
R:Shi, T.; Moriyoshi, K.; Sugihara, H.; Sakurada, K.; Kadotani, H.; Yokoi, M.; Akazaw
J. Biol. Chem. 269, 2836-2843, 1993
A:Title: Molecular characterization of the family of the N-methyl-D-aspartate receptor
A:Reference number: A45219; MUID:93155102; PMID:8428958
A:Accession: D45219
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1265-1323 <ISH>
A:Experimental source: rat
A>Note: sequence extracted from NCBI backbone (NCBIP:124265)
C:Superfamily: N-methyl-D-aspartate receptor 2D; glutamate receptor homology
F,451-879/Domain: glutamate receptor homology <GRH>

```

```

Query Match 2.8%; Score 112; DB 2; Length 1323;
Best Local Similarity 22.9%; Pred. No. 12; Mismatches 157; Indels 152; Gaps 28;
Matches 106; Conservative 48;
QY 162 INIALMSGGEALWORYLALGLTQALINEFFGPAFLAMGRKGNLH-----TWGPIUP 215
    |||
    |||
Db 48 LNVALVFSSP-----AYAAEARL-----GPAVAAVRRPGLDVRPVALVINGSDP 93
QY 216 PSMHIKQLYQ-----HRVL--DQWSPGMPVLPAPAGHYPEAVTRVFPQVNTV 263
    |||
    |||
Db 94 PS-----LVQLQDLDLSGLRGVHGVEFDDSRRAVAPILDFLSAQSLPIVAVHG-----143
QY 264 KMGSGWHFNCSSYCSFLAPEDDPIPTIGSLFIRELIKEFTGTHIYAGDTFENMQ-----318
    |||
    |||
Db 144 -----GAAVLTPKPK-----GSTFQ-----LGSSTEQQLQVIFEV 175
QY 319 -PSSPSRYAAATTA-----VYEAATVPTREAVTWLQGMFQHQPF--WGPQAIRAVLGA 372
    |||
    |||
Db 176 LEEYDWTSPFAVYTRAPGRAPFLSYIEVLTG--SLVGN--EHRGALLTDPGAGEAVLGA 231
QY 373 VPRG-----RLVLDLFAESQPYV--TRTASFOGPFIMCM--HNGGNGHGLFGALEAVN 424
    |||
    |||
Db 232 QLRVSVAQIRLLFC--AREEAPVFAEAGLGPVYVFWVQPLAGGG-----S 282
QY 425 GSEEARLFPNSTWGTMAREGISONEVYSLMELGWRKDPVPLDAMWTSFA-----479
    |||
    |||
Db 283 GVPGEPILLPFGSSPLPAG-----FAVASAGWMDLARRVVAAGVAVVAAQAQA 330
QY 480 -ARRYGSHPDGAAMRLLRSVYNSGEGACGHNRSPLVRRPRLQ--MNTSIWYNSD 535
    |||
    |||

```

Db 331 LIRDYGL-PEL-----GHDCRTQNRTH--RGSLSRYFNMTI-MDNRDY 371  
 Oy 536 VFEAWRLITTSAPSLATSPAFRYDLDTQAOVELSVLYEE 578  
 Db 372 SFNEDGLVNV-PSLV-----VISTRTDRTWEEVGSWEQ 404

## RESULT 13

H87316  
 hypothetical protein CC0545 [imported] - Caulobacter crescentus  
 C/Species: Caulobacter crescentus  
 C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C/Accession: H87316  
 R/Niemann, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
 J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A/Title: Complete Genome Sequence of Caulobacter crescentus.  
 A/Reference number: A87249; MUID:21173698; PMID:11259647  
 A/Accession: H87316  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1027 <STO>  
 A/Cross-references: GB:AE005673; NID:913421736; PIDN:AAK22532.1; GSPDB:GN00148  
 C/Genetics:  
 A/Gene: CC0545

Query Match 2.8%; Score 111; DB 2; Length 1027;  
 Best Local Similarity 22.5%; Pred. No. 10;  
 Matches 157; Conservative 47; Mismatches 217; Indels 278; Gaps 38;

Oy 135 VCTOSYFVW-----MDMNRERE-----IDMMLNGILAL-----AMSGEAIW 175  
 Db 68 VCDVSATGQWTLRASERAMDLKRLNDEGSLRKAIGMRRDGADPATADLDLIGKAFS 127  
 Oy 176 ORVYLATGLDQAEINERFTGPAPFLAGMGNGNLHTWDGFLPPSWH-IKOLYLQHRVIDDKR 234  
 Db 128 PKA-MAMMERITEQEAAGBEL---RLATALEMSGALPAPALLARLQAALNRDQAR 183  
 Oy 235 SFGMTPLVPAFAGHVPAAVTRVPPQVNVTKMGSHFNCSYSCSFLIAP-EDPI----- 287  
 Db 184 R-GELLADGFOGR-----EQAVETLTIDW-----INPDQDIFIRVAVY 221  
 Oy 288 ---PFIIG-SLFLBELIKEFTDHI-----YGADTNEM-----OPSSSE 324  
 Db 222 VSGI-PALGKSTLLHVRQVAAQCGEVTVVRLDPDRASLDVLQGLGSTEVAQVASSQ 281  
 Oy 325 SYLAATTAAYEANT-----GAYVRGILL-VYDLPAES-----AVDTEAVWL 348  
 Db 282 SAADLDQARLEAITTQETIRLKGGSGIPYALIGVQLQAGRLVWVLDTEVAR 341  
 Oy 349 LOGWLFQHQFQFQPAQIRAVL-----GAYVRGILL-VYDLPAES----- 387  
 Db 342 GRGETHPRQLFDMDQDLAPVAVLVVVAAGRGSLPVDRAVETLNDLTDAAERL 401  
 Oy 388 -----QPVYTRT-ASFGQGFIMCMLEHNGNHS-----LFGALEAVNGG 426  
 Db 402 LSADVPALMTQYIALAEGDPLQRLATLARREGAGLTKGKAGITTYRSLERLGG 461  
 Oy 427 PEA-ARLP-----NSTMGVTGMAP-----EGISQNEVYISLMAEL-----GW 463  
 Db 462 RHAKIARFIPLFRFNSSELLGAVMAPVVLGERLSPGAAA-ALVAMELDQDTWLAGEBEGW 520  
 Oy 464 RKDPVPD-----LAAWVVSFAARVGVSHPDAGAMRLLLSVYVCSGEACRGHNSPLVR 519  
 Db 521 -LTQPDRLRRLAGLYDEHARRRGALHRRR-AAW-----FAE 556  
 Oy 520 RPSIQMNTSIWNSDVFEAWRLITLSA---PSIATSPAR-VYLLDTQAOVELSVLY 575  
 Db 557 RPEIWMAGAESLYHR---LQATRWAGPDALKGMNLAAGVAFOSADLEELP----- 602  
 Oy 576 YEEBASAVLSKELASLIRAGGVLAWE-----LLP-----ALDEV-----LASDR 615

Db 603 -EABADA-----LIRARGERSYEGRGDALTPSGVSPAADELNMLISKGLAEASH 653  
 Oy 616 FL-----LGSWLEQARAAA 629  
 Db 654 FYGRVFPALDPYSPFDADVALLMRTGRWTEARRLLA 692

## RESULT 14

F87285  
 conserved hypothetical protein CC0295 [imported] - Caulobacter crescentus  
 C/Species: Caulobacter crescentus  
 C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C/Accession: F87285  
 R/Niemann, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
 J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A/Title: Complete Genome Sequence of Caulobacter crescentus.  
 A/Reference number: A87249; MUID:21173698; PMID:11259647  
 A/Accession: F87285  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-285 <STO>  
 A/Cross-references: GB:AE005673; NID:913421436; PIDN:AAK22282.1; GSPDB:GN00148  
 C/Genetics:  
 A/Gene: CC0295

Query Match 2.8%; Score 109.5; DB 2; Length 285;  
 Best Local Similarity 23.7%; Pred. No. 2.1;  
 Matches 80; Conservative 36; Mismatches 90; Indels 131; Gaps 17;

Oy 328 AATTAIVYEAHTAVDTEAVNLQGLMFLQHQFQFQPAQIRAVLAVAPRGRLVLDLFAES 387  
 Db 3 AAPTAAVYAGLTAWSLATTTCAMF-----ARRRLADGRVLSVE----- 43  
 Oy 388 QPVYTRTASFOGQPF-----IMCMLEHNGNHLFGALAE----- 422  
 Db 44 ---QSLAQGATYAAWLPVIGIWLVLRRFGA--GLRGVLAAPATGLVTPLEALVAS 96  
 Oy 422 ---VNGSGPAAR-----LFPNSTMGTCM-----AEGISQNEVYISLMAE- 460  
 Db 97 LIDQTFAGGGDLAEALGRAPVCILHATVAVGAAAHRRABASAKNTLLQALAE 156  
 Oy 461 -----LGRKDPVPDLAAWVTSFAARVGVSHPDAGAMRLLLSVYVCS 505  
 Db 157 RATSVDVAERLMWAGARRVFPV-DITAVEMFGAADNVVYHMGREG--LNRATLQ-S 211  
 Oy 506 GEA-----CRGNRSPVLRPSIQMNTSIWYNSDVFEA-----WRLLTSAPSLA 551  
 Db 212 LEARLDPRLEFARAH-RSALV-----MLSKVSOAQPLSDGSRLLTLASGAEV 257  
 Oy 552 TSPAFRYDLTD-LTRQAOVELSVLYEEARAGAYLSKE 587  
 Db 258 TSKRTYRDALIKRLGRQA-----SDSQAAHNSSE 285

## RESULT 15

T47641  
 hypothetical protein T15C9.20 - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C/Accession: T47641  
 R/Mewes, H.W.; Rudd, S.; Lemcke, K.; Meyer, K.F.X.  
 submitted to the Protein Sequence Database, April 2000  
 A/Reference number: Z24470  
 A/Accession: T47641  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-814 <NEW>  
 A/Cross-references: EMBL:AL122970  
 A/Experimental source: cultivar Columbia; BAC clone T15C9  
 C/Genetics:  
 A/Map position: 3



A: Introns: 28/2; 51/3; 252/3; 312/3; 350/3; 370/2; 392/3; 410/1; 441/1; 469/1; 604/3; 62  
A: Note: T1509.20

Query Match 2.8%; Score 109.5; DB 2; Length 814;

Best Local Similarity 20.6%; Pred. No. 9.3;  
Matches 113; Conservative 69; Mismatches 192; Indels 175; Gaps 26;

```
QY 150 WEREIDWALNGINIALAMSGOEAIWQRYVALGLTQAEINEFFTPAFLAMGRMYLAT 209
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 233 WKEELEVALIRGGVPWAL-----RGELMQAF--VGVRKRCKDYQ-----NLTA 274

QY 210 WDGPLPESMHIKQYLQHRVLDDKRSFCMTFVLDPAPAGHVEATRVPEQVNTVKGSMG 269
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 275 ADGSYNT---TEQEMQH-VDDKSSSTSIATVEKMKQIKEDLPRTFP-----G 320

QY 270 HFNCSYSCSFLAPEDPIFFIIGSLFLRELKEFGTDHIYGADTFNEMQPSSEPSYLA 329
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 321 H-----PALDDGRVALRRLLTAYARHN-----PSVGYC 349

QY 330 ATTAYEAMTAV---DTEAVMLQGMLEQHQPQFWGPAQITRAVLGAVPRGRLIVLDFAE 366
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 350 QAMNFFALLLMLPEENAFMALIGLIDYNGYISEMTESQV---DQLVLEELVRB 404

QY 387 SQPVYTRTASFQGFIFWCMLEHFGNHLFGALEAVNGSPEAARLPNSTWVGTGMAP- 445
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 405 RFPKLVHHLDYLGVCVAVT-----GFWFLIFMN-----MLPW 438

QY 446 EGISONEVYYSIMELGWRKDPVPDLAAMVTSFAARR-YG--VSHPDAGAMRL---LL 498
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 439 ESVLR---VMDVLLFEGTR-----VMLFTALIMELYGPALVTTKDGADAVTLQSLT 489

QY 499 RSVNCSG---EACGHRSPVLRPSLOMN-----TSIWTNRSDVFEAWR----- 541
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 480 GSTPSSQVLTLACMGYQNVHEIRLQELRSKRPAVIAALEERSKGLQAMRDSKGLASKL 549

QY 542 -----LLTSAPSLATSPAPRYD-----LLDLT-----RQAVQELVS 573
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 550 YNFKODPKSVLVDSKASLSNGSLSRSESGSSNADEVLSLTGDEGVDSVODLQAQAEICK 609

QY 574 LYIEARSAVYLSKELASLLRAGVLAAYELLPALDEVLASDSRFLGSMLEQ--ARAAYS 631
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 610 L-LERKSALLRAE-----ELETALMEIYKEDNRQLSAKVEQLBOEMAEVQ 655

QY 632 EAADFYEQ 640
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 656 RLSDKQEQ 664
```

Search completed: February 13, 2004, 16:16:27  
Job time : 37 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2004, 13:34:59 ; Search time 19 Seconds

(without alignments)  
1838.992 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939  
Sequence: 1 MEAVAVAAVGVLLIAGAG.....VDLAKTFLKTYPGWVAGSW 743

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_41.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3939	100.0	743	1	ANAG HUMAN
2	123.5	3.1	497	1	CPDH MACFA
3	119	3.0	497	1	ACVS NOCLA
4	117.5	3.0	497	1	CPD6 HUMAN
5	110.5	2.8	1029	1	YK35 RHIME
6	109.5	2.8	1377	1	RHSA_ECOLI
7	109.5	2.8	1411	1	RHSE_ECOLI
8	108	2.7	726	1	CATA_SALTY
9	104.5	2.6	722	1	CATA_ECOLI
10	103	2.6	698	1	YMCA_ECOLI
11	101	2.6	658	1	P85B RAT
12	101	2.6	3519	1	OL56_STRAT
13	100	2.5	726	1	CATA_SALTY
14	100	2.5	851	1	ENV_HV2DI
15	100	2.5	1323	1	NME4_MOUSE
16	100	2.5	1336	1	NME4_MOUSE
17	99.5	2.5	1092	1	LIFR_MOUSE
18	99	2.5	624	1	SIR_STYPT
19	99	2.5	1330	1	UCS4_SOLITU
20	99	2.5	1587	1	LMG3_HUMAN
21	98.5	2.5	512	1	NUOM_RHOCA
22	98.5	2.5	1774	1	MSAS_PENPA
23	98	2.5	860	1	ENV_HV2BE
24	98	2.5	973	1	VP18_HUMAN
25	98	2.5	1323	1	NME4_MOUSE
26	98	2.5	2067	1	BIMB_EMENT
27	97.5	2.5	497	1	GLPK_BACED
28	97.5	2.5	693	1	UL47_HSV11
29	97.5	2.5	786	1	AAC_ACTUT
30	97.5	2.5	842	1	VGLH_HSVBC
31	97.5	2.5	1155	1	C2TA_MOUSE
32	97	2.5	1094	1	EMBC_MYCTU
33	97	2.5	2594	1	7LES_DROVI

34	96.5	2.4	856	1	ENV_HV1MN	P05877 human immun
35	96	2.4	708	1	NICA_MOUSE	P57716 mus musculus
36	96	2.4	2314	1	PTP2_HUMAN	P23471 homo sapien
37	95.5	2.4	330	1	RCEM_ROSDE	P26279 roseobacter
38	95.5	2.4	914	1	PERT_RAT	P14650 rattus norv
39	95.5	2.4	970	1	ANDI_MOUSE	P59328 mus musculu
40	95	2.4	344	1	LICI_BURCE	P22089 burkholderi
41	95	2.4	636	1	GIDA_BRUME	Q8Y355 bruceella me
42	94.5	2.4	664	1	UL47_HSV1F	P08313 herpes simp
43	94.5	2.4	761	1	CTPA_MYCTU	Q10876 mycobacteri
44	94.5	2.4	866	1	NASA_KLEPN	Q06457 klebsiella
45	94	2.4	500	1	CPDG_CAVPO	Q64403 cavia porce

## ALIGNMENTS

RESULT 1  
ANAG\_HUMAN STANDARD; PRT; 743 AA.  
AC P54802;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Alpha-N-acetylglucosaminidase precursor (EC 3.2.1.50) (N-acetyl)-alpha-glucosaminidase (NAG).  
GN NAGLU OR UFGSD1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
[1]  
PP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=66372812; PubMed=8650226;  
RA Zhao H.G., Li H.H., Bach G., Schmidtchen A., Neufeld E.F.;  
RT "The molecular basis of Sanfilippo syndrome type B.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:6101-6105(1996).  
[2]  
PP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  
RX MEDLINE=66372812; PubMed=8776591;  
RA Weber B., Blanch L., Clements P.R., Scott H.S., Hopwood J.C.;  
RT "Cloning and expression of the gene involved in Sanfilippo B syndrome (mucopolysaccharidosis III B).";  
RL Hum. Mol. Genet. 5:771-777(1996).  
[3]  
PP SEQUENCE FROM N.A.  
RA Zhao Z., Yazdani A., Shen Y., Sun Z.S., Bailey J., Caskey C.T.,  
RX Lee C.C.;  
RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.  
[4]  
PP VARIANTS MPS-IIIB HIS-92; SER-115; CYS-140; LYS-153; LEU-358; VAL-664 AND ARG-682.  
RX MEDLINE=98107938; PubMed=9443878;  
RA Schmidtchen A., Greenberg D., Zhao H.G., Li H.H., Huang Y., Tieu P.,  
RA Zhao H.-Z., Cheng S., Zhao Z., Whitely C.B., di Natale P.,  
RX Neufeld E.F.;  
RL "NAGLU mutations underlying Sanfilippo syndrome type B.";  
RL Am. J. Hum. Genet. 62:64-69(1998).  
[5]  
PP VARIANTS MPS-IIIB.  
RX MEDLINE=99133861; PubMed=9950362;  
RA Bunge S., Knigge A., Steglich C., Kleijer W.J., van Diégelen O.P.,  
RA Beck M., Gal A.;  
RT "Mucopolysaccharidosis type IIIB (Sanfilippo B): identification of 18 novel alpha-N-acetylglucosaminidase gene mutations.";  
RL J. Med. Genet. 36:28-31(1999).  
[6]  
PP FUNCTION: INVOLVED IN THE DEGRADATION OF HEPARAN SULFATE.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-acetyl-D-glucosamine residues in N-acetyl-alpha-D-glucosaminides.  
CC -1- SUBUNIT: Monomer and homodimer.  
CC -1- TISSUE SPECIFICITY: LIVER, OVARY, PERIPHERAL BLOOD LEUCOCYTES, TESTIS, PROSTATE, SPLEEN, COLON, LUNG, PLACENTA AND KIDNEY.

CC -1- DISEASE: DEFECTS IN NAGLU ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS  
 CC TYPE IIIB (MPS-IIIB) (ALSO KNOWN AS SANFILIPPO B SYNDROME). AN  
 CC AUTOSOMAL RECESSIVE DISORDER WHOSE CLINICAL FEATURES ARE SEVERE  
 CC MENTAL DETECTIONATION BUT MILD SOMATIC MANIFESTATIONS IN CHILDHOOD,  
 CC AND DEATH IN THE SECOND DECADE. BIOCHEMICALLY, THIS DISEASE IS  
 CC CHARACTERIZED BY UNDEGRADED OR PARTIALLY DEGRADED HEPARAN SULFATE  
 CC WHICH ACCUMULATES IN LYSOSOMES AND IS EXCRETED IN URINE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; U43573; AAC50512.1; -;  
 DR EMBL; U43573; AAC50513.1; -;  
 DR EMBL; U40846; AAB06188.1; -;  
 DR EMBL; U78464; AAB36604.1; -;  
 DR PIR; G02270; G02270.  
 DR Genew; HGNC; 7632; NAGLU.  
 DR MIM; 252320; -;  
 DR GO; GO:0005764; C:lysosome; TAS.  
 DR GO; GO:0004561; F:alpha-N-acetylglucosaminidase activity; TAS.  
 DR GO; GO:0007399; P:neurogenesis; TAS.  
 DR Pfam; PF05089; NAGLU; 1.  
 DR Hydrolase; Glycosidase; Glycoprotein; Mucopolysaccharidosis;  
 KM Signal; Disease mutation; Polymorphism.  
 FT CHAIN 1 23  
 FT 24 743  
 FT 59 743  
 FT 68 71  
 FT 84 87  
 FT 134 134  
 FT 261 261  
 FT 272 272  
 FT 435 435  
 FT 503 503  
 FT 526 526  
 FT 532 532  
 FT 79 79  
 FT 92 92  
 FT 100 100  
 FT 115 115  
 FT 140 140  
 FT 142 142  
 FT 153 153  
 FT 243 243  
 FT 277 277  
 FT 280 280  
 FT 292 292  
 FT 358 358  
 FT 452 452  
 FT 482 482  
 FT 561 561  
 FT 561 561

FT VARIANT 565 565 R -> Q (IN MPS-IIIB).  
 FT /FtId=VAR\_008989.  
 FT R -> H (IN MPS-IIIB).  
 FT /FtId=VAR\_005012.  
 FT A -> V (IN MPS-IIIB).  
 FT /FtId=VAR\_005013.  
 FT R -> H (IN MPS-IIIB).  
 FT /FtId=VAR\_005014.  
 FT L -> R (IN MPS-IIIB).  
 FT /FtId=VAR\_005015.  
 FT E -> K (IN MPS-IIIB).  
 FT /FtId=VAR\_008990.  
 FT G -> R.  
 FT /FtId=VAR\_008991.  
 FT A -> L (IN REF. 2; AA SEQUENCE).  
 FT S -> L (IN REF. 2; AA SEQUENCE).  
 FT S -> L (IN REF. 2; AA SEQUENCE).  
 SQ SEQUENCE 743 AA; 82166 MW; 6D8D6A42C7BA6083 CRC64;  
 Query Match 100.0%; Score 3939; DB 1; Length 743;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-277;  
 Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MEAAVAAAVGVTLLAGAGAGDEAREAAVRLVARTLGGPADPVSVERALAKP 60  
 DB 1 MEAAVAAAVGVTLLAGAGAGDEAREAAVRLVARTLGGPADPVSVERALAKP 60  
 QY 61 GLDTYSLGGGGAARVVRGSGTGAAGLHRYLDFGCGHVAWSGQLRPLRPVAVGE 120  
 DB 61 GLDTYSLGGGGAARVVRGSGTGAAGLHRYLDFGCGHVAWSGQLRPLRPVAVGE 120  
 QY 121 LTEATPNRYRYQVNCVCSFVWMDVAREERIDMMLNGINLALMSGEALWQRYVL 180  
 DB 121 LTEATPNRYRYQVNCVCSFVWMDVAREERIDMMLNGINLALMSGEALWQRYVL 180  
 QY 181 ALGTLQAEINEFFGPAPFLAMGRGNLHTWGPPLPSPMHIKQYLQHRVLDQMSFCWTP 240  
 DB 181 ALGTLQAEINEFFGPAPFLAMGRGNLHTWGPPLPSPMHIKQYLQHRVLDQMSFCWTP 240  
 QY 241 VLPAPAGHPVAVTRVPQVNTVMGSGHFNCSGCSFLIAPDPPIFPIIGSLFLELI 300  
 DB 241 VLPAPAGHPVAVTRVPQVNTVMGSGHFNCSGCSFLIAPDPPIFPIIGSLFLELI 300  
 QY 301 KEFGDTHYGADTENEQPPSEPSYLAATTAYEAMTAVDTAAWLLQWLFOHOPQF 360  
 DB 301 KEFGDTHYGADTENEQPPSEPSYLAATTAYEAMTAVDTAAWLLQWLFOHOPQF 360  
 QY 361 WGPAPQIRAVLGAVERGRLLVLDLFAESQPYVTRTASFQGPFIWCMHNFQGNHFLGAL 420  
 DB 361 WGPAPQIRAVLGAVERGRLLVLDLFAESQPYVTRTASFQGPFIWCMHNFQGNHFLGAL 420  
 QY 421 EAVNGGPEAAALFENSTMTVGMAPEGISONEVYYSIMAEIGMRKDPVPLAAVTSFPA 480  
 DB 421 EAVNGGPEAAALFENSTMTVGMAPEGISONEVYYSIMAEIGMRKDPVPLAAVTSFPA 480  
 QY 481 RRYGVSHDAGAAVRLILRSYVNGSGEACRGNHNSPLVRBPSLOMNTSIWNRSDVEAW 540  
 DB 481 RRYGVSHDAGAAVRLILRSYVNGSGEACRGNHNSPLVRBPSLOMNTSIWNRSDVEAW 540  
 QY 541 RLITTSAPSLATSAFRYDLDLTLROAVOEIVSYIYEASAYISKELASLLRAGGLAY 600  
 DB 541 RLITTSAPSLATSAFRYDLDLTLROAVOEIVSYIYEASAYISKELASLLRAGGLAY 600  
 QY 601 ELPLALDEVILASDSRFLGSLLEBARAAVSEADPFEONSRQYLTLMGPEGNIIDYAN 660  
 DB 601 ELPLALDEVILASDSRFLGSLLEBARAAVSEADPFEONSRQYLTLMGPEGNIIDYAN 660  
 QY 661 KQLAGLVANVTPKWRFLFLEALVDSVAQGIPOQHOPDKVVFQLEQFVLSKQYPSQPR 720  
 DB 661 KQLAGLVANVTPKWRFLFLEALVDSVAQGIPOQHOPDKVVFQLEQFVLSKQYPSQPR 720  
 QY 721 GDTYDLAKKIFLKXYPGWVAGSW 743  
 DB 721 GDTYDLAKKIFLKXYPGWVAGSW 743

```

RESULT 2
CPDH_MACFA
ID_CPDH_MACFA STANDARD, PRT, 497 AA.
AC 02948;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 2D17 (EC 1.14.14.1) (CYIID17).
GN CYP2D17
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OC NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Lawton M.P., Laddison K.J., Speirs A.A., Mankowski D.C., Tweedie D.J.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1 CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -1 SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC similarity).
CC -1 SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38218; AAA79722.1; -
DR PIR; G02938; G02938.
DR HSSP; P00179; 1DT6.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450_1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum.
FT METAL 443 443 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SO SEQUENCE 497 AA; 56010 MW; 3594AA88F0458B1 CRC64;

Query Match 3.1%; Score 123.5; DB 1; Length 497;
Best Local Similarity 22.9%; Pred. No. 0.21;
Matches 104; Conservative 47; Mismatches 140; Indels 163; Gaps 22;

343 TEAVWL-----OGWLFQHOPOFMGPAQIRAVLGVPRGRLLVLDLFABSQPVYTRT 394
12 TVALFLVLVDLMHRRORWARIYP--GDLPLPGL-----GULLHYD----- 50
395 ASFOGQPIWCLMNFQGN-----HGLFGALFA-VNGSPEAART--FPNST 437
51 --FKMTPYCFDPLRRRFGNVFSLQIAMPVYVNLGLAAVEALVTCEDPTADRPVYINQ 108
438 MVGTGMAPEGI-----SQHEVYVSLMAELGWRKDPVLDLAAVTSPFAARYVSH 487
109 VLGQPRSQGVFLARYGPAMREQRFSVSTIRNLGLGK--KSLEQWTEBAACLCAPT 165
488 PDAGAAAR--LLRSVYVNCGEACRGHNSPLVRBSLQWNTSIWYNSDVPFAMRL 544
166 DQAGPFPRNSLLDKAVSN-----VIALTLTGR----- 193
545 TSAASLATSAPFRD-----LLDITROAVQELVSLYEEKRSVLSKELAS---LIR-- 593
194 -----RFEYDDPRFLRLFDLTLEALKE-----ESGFLRELVNALPILLRIP 234
594 --AGVLAVE--LLPALDEVL-----ASDRSFLIGSWLQPARAAVSEADPEYEQ 640

```

```

DB 235 GLAGKLRSGAKAFITQDELLETRMTWDPAPQPRDLTEAFLAEMEKAK-GNPSSFNEE 293
QY 641 NSRYQLTLMGPEGNITLDYANQLAGLVANYYPERRRFLFEALVDVSAQIPQCHQFDKN 700
DB 294 NLR-----MVVADLFSGMVTSTTLANGLLMLHPVQRRV---QGEIDV 338
QY 701 VFQLEQAFVLSKQRYP-----SQPRGDTVDL 726
DB 339 IQGVRRPEMGDQARMPYTTAVIVHEVQRFGDIVPL 372

RESULT 3
ACVS_NOCCLA
ID ACVS_NOCCLA STANDARD, PRT, 3649 AA.
AC P27743;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE N-(5-amino-5-carboxypentanoyl)-L-cysteine]-D-valine synthase
DE (EC 6.3.2.26) (Delta-(L-alpha-aminoadipyl)-L-cysteine]-D-valine
DE synthetase) (ACV synthetase) (ACVS).
GN PCBAB.
OS Noccardia lactamdurans.
OC Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;
OC Pseudonocardineae; Pseudonocardiaceae; Amycolatops.
OC NCBI_TaxID=1913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VAR LC 411;
RX MEDLINE=92065808; PubMed=1956290;
RA Coque J.J.R., Martin J.F., Calzada J.G., Liras P.;
RT "The cephamycin biosynthetic genes pcbAB, encoding a large
RT multidomain peptide synthetase, and pcbC of Noccardia lactamdurans are
RT clustered together in an organization different from the same genes
RT in Acetomonas chrysoeum and Penicillium chrysogenum."
RL Mol. Microbiol. 5:1125-1133(1991).
CC -1 FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE
CC ACV ARE ACTIVATED AS AMINOACYL-ADENYLYATES WITH PEPTIDE BONDS
CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
CC INTERMEDIATES.
CC -1 CATALYTIC ACTIVITY: L-2-aminohexanoate + L-cysteine + L-valine
CC + 3 ATP = N-[L-5-amino-5-carboxypentanoyl]-L-cysteine]-D-valine +
CC 3 AMP + 3 diphosphate.
CC -1 COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETINES.
CC -1 PATHWAY: Biosynthesis of penicillin and cephalosporin, first step.
CC -1 SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -1 SIMILARITY: Contains 3 acyl carrier domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57310; CAA40561.1; -
DR PIR; S18268; S18268.
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR006163; PP_bind.
DR InterPro; IPR006162; Pantate_attach.
DR InterPro; IPR000379; Ser_estr_site.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 3.
DR Pfam; PF00568; Condensation; 3.
DR Pfam; PF00550; PP-binding; 3.
DR Pfam; PF00975; Thioesterase; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; 2.
DR PROSITE; PS00455; AMP BINDING; 1.
DR PROSITE; PS00075; ACP_DOMAIN; 3.

```

Ligase; Antibiotic biosynthesis; Multifunctional enzyme;  
 KM Repeat; Phosphopantetheine.  
 FT REPEAT 401 861 DOMAIN 1 (ADDPATE-ACTIVATING).  
 FT REPEAT 1014 1937 DOMAIN 2 (CYSTEINE-ACTIVATING).  
 FT REPEAT 2079 2985 DOMAIN 3 (VALINE-ACTIVATING).  
 FT DOMAIN 788 857 ACYL CARRIER (ACP) 1.  
 FT DOMAIN 1864 1933 ACYL CARRIER (ACP) 2.  
 FT DOMAIN 2910 2981 ACYL CARRIER (ACP) 3.  
 FT BINDING 820 820 PHOSPHOPANTHEINE (BY SIMILARITY).  
 FT BINDING 1896 1896 PHOSPHOPANTHEINE (BY SIMILARITY).  
 FT BINDING 2944 2944 PHOSPHOPANTHEINE (BY SIMILARITY).  
 FT ACT SITE 3502 3502 THIOESTERASE (BY SIMILARITY).  
 SQ SEQUENCE 3649 AA; 404079 MW; 6FD095704F858EB8 CRC64;  
 Query Match 3.0%; Score 119; DB 1; Length 3649;  
 Best Local Similarity 21.1%; Pred. No. 6.3; Indels 306; Gaps 37;  
 Matches 174; Conservative 63; Mismatches 283;  
 16 AGAGGAGDEAREAAVRAV-----ARLLGPAADPSVVERALAKPGDLTVSLGG 69  
 1988 AALPGALGTLVRHHPALRTLLKTDGVRQYIPADVRLEVP----- 2031  
 70 GGAARVVRGSGTGAAGAHLRLRDCGCHVAMSGQLRPRPFAVPGELTEATPNRY 129  
 2032 -----STVDSFAELDEVLT-----RAGVTRLEHLP-----RAAFDHD 2070  
 130 RYQVNTOSYSFVWMDMAREREI-----DMNLNGINIALAMSGEATVQRY 179  
 2071 EYLSVHVHSCFGWMDIFRELAALDGVPEALIGALRG-----YGEFAVWQRY 2124  
 180 LALGTOAINEPFTGAFIANGMGNLHT-----WGPDLPSNHR-----QYLCRVLD 231  
 2125 LT-GKRLAALTEWT-----GALGGFETIALPLDHPRPDYRGRELEFELDERTE 2176  
 232 QMRSGMT----- 239  
 2177 ALRLAATAVSLYSVLGAMCLMANYTGQHDLVVTEPANKGREPRAVGFANLLA 2236  
 240 -----PVLPAFHVPAVTRVPPQVNTKMGKGFNCYSGSFLIAPDP-----FPI 290  
 2237 LRVAVDPAATLPAVRSVGEAV-----VAAQVH-----GELPFOVYKELVEEDPSRHHI 2287  
 291 IGSFLFELIKFETDIYGAFTENEMPSSEPSYLAATVAEAMTVDETAWMLQ 350  
 2288 LQNLFTLQNV-----SDH-----TSALTGYQPS-----GMITTKFD-LSMTMETATGLA 2233  
 351 GWT-----FOHQPFQWGPQAQIRAVLGAVPGRLLVLDLFAESQ-PYV 391  
 2334 GNLTYAASLPDDTSAGFIATFKHV-----LAEFASAACTPIAQLTALDEPQALPDA 2388  
 392 TRTASFOGQPIWCMANFGNHLFGALEVNGGPEAALFNSMTVGTG---MAPEGI 448  
 2389 TRRARPPGCP-----GRCTRLP-----BEVAATWPRVAVVGDVRLTYREL 2430  
 449 SQ-----NEVVYSIMAEIGMKDP-----VPDLAAVMTSFAARVGVSHPDAGAA 493  
 2431 NERANRLAHHLRSVAERADELIALVLDKSELTVAILAVWKAAGAAVMPIDPSYDDRIA 2490  
 494 W-----RLLRSVNCSGACGAGHNRSPVLRPSLQMTSTWNRSDVFEAMRLILIS 546  
 2491 FMSDGTAKVL-----AGEAGHSRVRG-----LISGDVLDLBDQLDLTG 2529  
 547 APSLATSPAFRYDLDLTRCAVQELVSLYVEARSAVLSKELASILRAGVLAVELPAL 606  
 2530 EP-AENP-----VLETSTELAVAIY-----TSGTGPKVAVLHSGV----- 2567  
 607 DEVTASDSRFLGSLWLEQAPAAVSEADPYEQNSRYQLTIMGPEANIL----- 656  
 2568 DSFPAQLSGRYFGSPDSAEAVLFL-----ANYVDFSVQALSVLGHKLIVPPPSAAD 2624  
 657 -----DYANKOLAGLVANYTFRMLFLFALVDVSAQGIFFQOHFD 698  
 2625 PAFELANRE--GL-----SYLSTGTPQVERFD 2650

RESULT 4  
 ID CPD6\_HUMAN STANDARD; PRT; 497 AA.  
 AC P10635; Q16752;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cytochrome P450 2D6 (EC 1.14.14.1) (CYP1D6) (P450-DB1) (Debrisoquine 4-hydroxylase).  
 GN CYP2D6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88314109; PubMed=3410476;  
 RA Gonzalez F.J., Vilbois F., Hardwick J.P., McBride O.W., Nebert D.W., Gelboin H.V., Meyer U.A.;  
 RT "Human debrisoquine 4-hydroxylase (P4501D1): cDNA and deduced amino acid sequence and assignment of the CYP2D locus to chromosome 22.";  
 RL Genomics 2:174-179(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=88122614; PubMed=3123997;  
 RA Gonzalez F.J., Skoda R.C., Kimura S., Umeno M., Zanger U.M., Nebert D.W., Gelboin H.V., Hardwick J.P., Meyer U.A.;  
 RT "Characterization of the common genetic defect in humans deficient in debrisoquine metabolism.";  
 RL Nature 331:442-446(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90072069; PubMed=2574001;  
 RA Kimura S., Umeno M., Skoda R.C., Meyer U.A., Gonzalez F.J.;  
 RT "The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene.";  
 RL Am. J. Hum. Genet. 45:889-904(1989).  
 RN [4]  
 RP VARIANT LYS-281 DEL (CYP2D6\*9).  
 RX MEDLINE=93244880; PubMed=1844820;  
 RA Tyndale R., Aoyama T., Broly F., Matsumaga T., Inaba T., Kalow W., Gelboin H.V., Meyer U.A., Gonzalez F.J.;  
 RT "Identification of a new variant CYP2D6 allele lacking the codon encoding Lys-281: possible association with the poor metabolizer phenotype.";  
 RL Pharmacogenetics 1:26-32(1991).  
 RN [5]  
 RP VARIANTS SER-34 AND THR-486 (CYP2D6\*10).  
 RX MEDLINE=94115362; PubMed=8287064;  
 RA Yokota H., Tamura S., Puruya H., Kimura S., Watanabe M., Kanazawa I., Kondo I., Gonzalez F.J.;  
 RT "Evidence for a new variant CYP2D6 allele CYP2D6J in a Japanese population associated with lower in vivo rates of sparteine metabolism.";  
 RL Pharmacogenetics 3:256-263(1993).  
 RN [6]  
 RP VARIANT PRO-324 (CYP2D6\*7).  
 RX MEDLINE=95147995; PubMed=7845481;  
 RA Evert B., Griese B.U., Bichelbaum M.;  
 RT "A missense mutation in exon 6 of the CYP2D6 gene leading to a histidine 324 to proline exchange is associated with the poor metabolizer phenotype of sparteine.";  
 RL Naunyn-Schmiedeberg Arch. Pharmacol. 350:434-439(1994).  
 RN [7]  
 RP VARIANT GLU-212 (CYP2D6\*6B/6C).  
 RX MEDLINE=95172594; PubMed=7868129;  
 RA Daly A.K., Leathart J.B., London S.U., Idle J.R.;  
 RT "An inactive cytochrome P450 CYP2D6 allele containing a deletion and a base substitution.";

Hum. Genet. 95:337-341(1995).  
[8]  
RN VARIANT IIE-107 (CYP2D6\*17).  
RP MAESTRINI D., PERSSON L., BERTELSSON L., HAAELER J.,  
RX MASTIMIREMWA C., PERSSON L., BERTELSSON L., HAAELER J.,  
RA INGELMAN-SUNDERBERG M.,  
RT "A novel mutant variant of the CYP2D6 gene (CYP2D6\*17) common in a  
RL black African population: association with diminished debrisoquine  
RM hydroxylase activity".  
RN Br. J. Clin. Pharmacol. 42:713-719(1996).  
[9]  
RN VARIANT ARG-42 (CYP2D6\*12).  
RP MEDLINE=96209916; PubMed=8655150;  
RX MAREZ D., LEGRAND M., SABBAGH N.; Lo-Guidice J.M., Boone P., Broly F.,  
RA "an additional allelic variant of the CYP2D6 gene causing impaired  
RT metabolism of sparteine".  
RL Hum. Genet. 97:668-670(1996).  
[10]  
RN VNANTS.  
RP MEDLINE=97385645; PubMed=9241659;  
RX MAREZ D., LEGRAND M., SABBAGH N.; Guidice J.M., Spire C.,  
RA Lafitte J.J., Meyer U.A., Broly F.,  
RT "polymorphism of the cytochrome P450 CYP2D6 gene in a European  
RL population: characterization of 48 mutations and 53 alleles, their  
RM frequencies and evolution".  
RN Pharmacogenetics 7:193-202(1997).  
[11]  
RN VARIANT ARG-169 (CYP2D6\*14).  
RP MEDLINE=99164054; PubMed=10064570;  
RX WANG S.L., LAI M.D., HUANG J.D.;  
RA "G169R mutation diminishes the metabolic activity of CYP2D6 in  
RT Chinese".  
RL Drug Metab. Dispos. 27:385-388(1999).  
CC -1- FUNCTION: RESPONSIBLE FOR THE METABOLISM OF MANY DRUGS AND  
CC ENVIRONMENTAL CHEMICALS THAT IT OXIDIZES. IT IS INVOLVED IN THE  
CC METABOLISM OF DRUGS SUCH AS ANTIRHEUMATICS, ADRENORECEPTOR  
CC ANTAGONISTS, AND TRICYCLIC ANTIDEPRESSANTS.  
CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +  
CC oxidized flavoprotein + H(2)O.  
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.  
CC -1- INDUCTION: By pregnancy.  
CC -1- POLYMORPHISM: HIGHLY POLYMORPHIC. OXIDATIVE DRUG METABOLISM BY  
CC CYP2D6 IS CHARACTERIZED BY TWO PHENOTYPES, 'THE EXTENSIVE  
CC METABOLIZER (EM) AND POOR METABOLIZER (PM). OF THE CAUCASIAN  
CC POPULATIONS OF EUROPE AND NORTH AMERICA, 5%-10% ARE OF THE PM  
CC PHENOTYPE AND ARE UNABLE TO METABOLIZE THE ANTIHYPERTENSITIVE DRUG  
CC DEBRISOQUINE AND NUMEROUS OTHER DRUGS.  
CC -1- POLYMORPHISM: CYP2D6\*7 WAS ALSO KNOWN AS CYP2D6E, CYP2D6\*9 AS  
CC CYP2D6C, CYP2D6\*10 AS CYP2D6J, CYP2D6\*11 AS CYP2D6Z.  
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.  
CC -1- DATABASE: NAME=Cytochrome P450 Allele Nomenclature Committee;  
CC NOTE=CYP2D6 alleles;  
CC WWW="http://www.imm.ki.se/cypalleles/cyp2d6.htm".

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation --  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL, M20403; AA52153.1; -  
EMBL, X08006; CA30807.1; -  
EMBL, M33388; AA53500.1; -  
PIR, G01199; O4HUD1.  
HSSP, P00179; 1DT6.  
DR GENEW, HGNC:2625; CYP2D6.  
MIM, 124030; -  
INTERPRO: IPR001128; Cytochrome\_P450.  
PFAM, PF00067; P450\_1.  
PRINTS, PR00385; P450.  
PROSITE, PS00086; CYTOCHROME\_P450\_1.

KW	Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW	Mitochondrion; Endoplasmic reticulum; Polymorphism.
FT	IRON (HEME AXIAL LIGAND).
FT	V -> M (IN ALLELE CYP2D6*35).
FT	/FtId=VAR_008366.
FT	R -> H (IN ALLELE CYP2D6*21).
FT	/FtId=VAR_008367.
FT	R -> C (IN ALLELE CYP2D6*22).
FT	/FtId=VAR_008368.
FT	P -> S (IN ALLELE CYP2D6*10 and allele
FT	CYP2D6*14; poor debrisoquine metabolism).
FT	/FtId=VAR_008336.
FT	G -> R (IN ALLELE CYP2D6*12; IMPAIRED
FT	METABOLISM OF SPARTEINE).
FT	/FtId=VAR_001256.
FT	A -> V (IN ALLELE CYP2D6*23).
FT	/FtId=VAR_008369.
FT	T -> I (IN ALLELE CYP2D6*17; POOR
FT	DEBRISOQUONE METABOLISM).
FT	/FtId=VAR_008337.
FT	G -> R (IN ALLELE CYP2D6*14; POOR
FT	DEBRISOQUONE METABOLISM).
FT	/FtId=VAR_008338.
FT	G -> E (IN ALLELE CYP2D6*6B and allele
FT	CYP2D6*6C).
FT	/FtId=VAR_008339.
FT	A -> S (IN ALLELE CYP2D6*33).
FT	/FtId=VAR_008370.
FT	MISSING (IN ALLELE CYP2D6*9).
FT	/FtId=VAR_008347.
FT	R -> C (IN ALLELE CYP2D6*2, allele
FT	CYP2D6*12, allele CYP2D6*14 and allele
FT	CYP2D6*17; dbSNP:16947).
FT	/FtId=VAR_008340.
FT	I -> L (IN ALLELE CYP2D6*24).
FT	/FtId=VAR_008371.
FT	S -> L (IN dbSNP:1800754).
FT	/FtId=VAR_014633.
FT	H -> P (IN ALLELE CYP2D6*7; LOSS OF
FT	ACTIVITY).
FT	/FtId=VAR_008348.
FT	R -> G (IN ALLELE CYP2D6*25).
FT	/FtId=VAR_008372.
FT	I -> T (IN ALLELE CYP2D6*26).
FT	/FtId=VAR_008373.
FT	E -> K (IN ALLELE CYP2D6*27).
FT	/FtId=VAR_008374.
FT	S -> T (IN ALLELE CYP2D6*2, allele
FT	CYP2D6*10, allele CYP2D6*12, allele
FT	CYP2D6*14 and allele CYP2D6*17; impaired
FT	metabolism of sparteine).
FT	/FtId=VAR_008341.
FT	M -> V (IN REF. 3).
SO	SEQUENCE 374 AA; 374 MM; 543PADSFODEBCDAC CRC64;
QY	Query Match 3.0%; Score 117.5; DB 1; Length 497;
QY	Best Local Similarity 21.7%; Pred. No. 0.56; Indels 153; Gaps 22;
QY	Matches 97; Conservative 53; Mismatches 144;
Dy	345 AVML-----QGWLFCQPQFWGACRAVAGAVPRGRLLVLDFAEASQPVYTRTAS 396
Dy	14 AIFLLVDLMHRBRGMARAYPP--GLPLPGL-----GNLHVD----- 50
Dy	397 FQGGPFWMCLAHNFGGN-----HGFLGALHA-VNGSPFAARL--EPNSTMV 439
Dy	51 FQNTPEFYFDQLRRFRFGDVFSLOLTAMTEVVVNLGAIAVEALVTHGEPTADRPVPVITQIL 110
Dy	440 GTGMAPEGI-----SQNEVVSILMAELGRKDKPVEDLIAMVTSFAARRGVSHPD 489
Dy	111 GFPGRSQGVTLARYGPAMRQRARRSVSTLNLTGKG--KSLEGMVIEEA----- 158
Dy	490 AGAAMRLILASVYNGSGEACRGHNRSPLVRBPISQMTSITWYNSDVFANRLILLTSAPS 549

Db 159 -----CLCAFANNSGRPF--RPNGLDKAV---SNVI-----AS 188  
 QY 550 LATSPARRYD-----LLDTTRCAVQELVSLYEASAYLSKELASLIPAGVLAWE-- 601  
 Db 189 LTCRRREYDDPRFLRLILDLAQDEGLKESGFLREVLNAVPLILHPL--AGVTLRQKA 246  
 QY 602 LPLPDLVYL-----ASDPRFLGSMLEQAPAAAVSEADPFYQNSRYQLTLWGPE 652  
 Db 247 FLTQIDELITLHRTWDPAPQPPDLTEFLAEMERAK-GNPESSEFNDENLR----- 296  
 QY 653 GNILDYANKOLAGVANYVTTPRMFLFEALVDVAQGFPGQOFDKNVQLER----- 706  
 Db 297 ---IVADLFSAQGVTTSTTLAKGLMLTHPDVQRV---QGEIDVITQVRRPENGQ 350  
 QY 707 -----AFVLSKQRYPSQPRGDTVDL 726  
 Db 351 AHMPYTFVHIEVQRF-----GDIVPL 372

RESULT 5  
 YK95\_RHIME STANDARD; PRT; 1029 AA.  
 ID YK95\_RHIME  
 AC Q52999;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein R02095.  
 GN R02095 OR SKC01491.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 CC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
 CX NCBI\_Taxid=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; Pubmed=11481430;  
 RA Capela D., Barloy-Hublier F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Boistard P., Becker A., Boutry M., Cadieu B., Dreano S., Gloux S.,  
 RA Godrie T., Goffeau A., Kahn D., Kise E., Lelaurie V., Maury D.,  
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampegege U.,  
 RA Renard C., Thebaud F., Vandenbol M., Weidner S., Galibert F.,  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 RT Sinorhizobium meliloti strain 1021.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 RN [2]  
 RP SEQUENCE OF 1-483 FROM N.A.  
 RC STRAIN=1021;  
 RA Bent A.F., Signer E.R.,  
 RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 1 guanylate cyclase domain.  
 CC -1- SIMILARITY: TO M.LEPRAE MLCB2407.09.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO  
 CC FRAMESHIFTS IN POSITIONS 311 AND 321.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 CC EMBL; AL591789; CAC46674.1;  
 DR EMBL; M30934; AAA88525.1; ALT FRAME.  
 DR InterPro; IPR001054; G\_cyclase.  
 DR SMART; SMO0044; CYC; 1.  
 DR PROSITE; PS50125; GUANYLATE CYCLASES 2; 1.  
 KM Hypothetical protein; ATP-binding; complete proteome.  
 FT DOMAIN 39 168  
 FT NP\_BIND 261 268 GUANYLATE CYCLASE.  
 FT CONFLICT 323 323 A -> P (IN REF. 2).  
 SQ SEQUENCE 1029 AA; 112709 MW; 4E4A78188B601958 CRC64;

Query Match 2.8%; Score 110.5; DB 1; Length 1029;  
 Best Local Similarity 19.0%; Pred. No. 4.8;  
 Matches 169; Conservative 98; Mismatches 279; Indels 343; Gaps 42;

QY 17 GAGGAAGDEAREAAVRLVLRLLGPGF-----AADSVSEVERLAAPGD 63  
 Db 26 GGSSTVGGERR---IVTALCYDVLVSTDLNVMYDIEDYQELMSAFQLSKQALASHSGVM 82  
 QY 64 TYSLOGGGA---RVRVSGTGVAAGAAHRYLDFC----- 97  
 Db 83 QHEADGGVALPEPIELKKAASLAIRAGLG---IVEACKRVAREAGQDDLVRVGIATSV 140  
 QY 98 -----GCHYAM-----SSQQLRPPRLPAV--PGLTATNRYKYVNCVQGSFPMW 145  
 Db 141 ALVLEGSREGWREPEVTAAALMAARLQATIPNSVLSEETR---LGRSTAFVF- 194  
 QY 146 DWARWEREIDWALNGINILALAMSGQEAIMQVYIALGLTOAEINEPFGPAFLAMG 205  
 Db 195 -----QGSKEIKGFAPBEKVWALGHKVGVD-----FYAFGRMG 229  
 QY 206 -----NLHT---WDGPLEPSWHIKQLYLQ-----HRVLDQNR----- 234  
 Db 230 GELINRENEINLTGQLMDGVLAGGSV---VLIGDAGIGKSRLLREIRTRAKSKLLF 287  
 QY 235 -----SFGMTPLYFAPAGHVEPAVTRVPQVNVTKKSGHFNCSYSCSFLIAPEDPIPI 290  
 Db 288 PCLTEGGRSTHPLPLNLPRAYS-----GSAQGMG-----PTAA 323  
 QY 291 IGSFLRELIEKFGT---DHLYGADTFENQPPSSPSYLAATTAVYEMTAVDTAV 346  
 Db 324 VALTEFERNIGIRPAVDVFSYLLAGQ--SRLOQMEDPFAIEKARFALRLALEACRR- 380  
 QY 347 WLQGMQLGHQPGFWGPAQI-----RANLGAVR--GRLVLDLFAESQOPY- 391  
 Db 381 -----GPAVVAVEDVHMIDPISRDLSGEAAI IAKFPVL-LYTSRSBYA 424  
 QY 392 -----TRTASFOGQPIWCMHFNQNGHFGALBAVNGPE-- 428  
 Db 425 SEMLDAAPTRLALPLDSDERLAIKAWP---EHLALPLPLFATERISGVPLF 479  
 QY 429 -----AARLPNSTWVGCMABEGISQNE-VVYSIMALQMRKD-----P 467  
 Db 480 IEIICQWVSQVPEPTMELSEBANSVSAFESIESRLQDLGTRERVAAAAVAGTGT 539  
 QY 468 VPDLAAMV-----TSPFAAR-----RYGVSHPDAGAAWRLLSRYVN 503  
 Db 540 LPLRALPLDPGKSLANAAADLTCTGFLTRIVRGRTAYGFRH-----TLIGETIYN 592  
 QY 504 CSGEACRGHNRSPVLRPSLQNTSITVNSDVFPAWFLILTSAPSLATSPAFRYDLIDL 563  
 Db 593 -----AVLRKQOVHLHR---LFTAVNQNRG--MAAW-----IDT 622  
 QY 564 TQAVQVELVSLYEASAYLSKELASLIPAGVLAWEYLLPDLDEVLASDGRFLGSMLE 623  
 Db 623 GALA-----EHAERAGLVEAVPLFIAG-----KESSRSAMIEROLE 663  
 QY 624 QAPAAVSEADPFYQNSRYQLTLWGPEGNILDYANKOLAGLVANYTTPRMFLFEALV 683  
 Db 664 HALDLCQGSSEDDTALXKLALTLALGP-----ILIGVGLSSEPARRLYEDAV- 712  
 QY 684 DSVAGGIP--QGHQF-----DKNYFQ--LEQAVLSQGRPE 716  
 Db 713 -DIARRRPSQSQWFIYWGWLTLGQDFRVMDRALVRSLSKANE 760

RESULT 6  
 ID RHSA\_ECOLI STANDARD; PRT; 1377 AA.  
 AC P16916;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE RhsA protein precursor.



GN RHSA OR B3593.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxId=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90094253; PubMed=2403547;  
 RA Feulner G., Gray J.A., Kirschman J.A., Lehner A.F., Sadosky A.B.,  
 RA Vlazny D.A., Zhang J., Zhao S., Hill C.W.;  
 RT "Structure of the rhaA locus from Escherichia coli K-12 and  
 RT comparison of rhaA with other members of the rha multigene family.";  
 RL J. Bacteriol. 172:446-456(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94316500; PubMed=8041620;  
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
 RT region from 76.0 to 81.5 minutes.";  
 RL Nucleic Acids Res. 22:2576-2586(1994).  
 RN [3]  
 RP REVIEW.  
 RX MEDLINE=95020608; PubMed=7934896;  
 RA Hill C.W., Sandt C.H., Vlazny D.A.;  
 RT "rha elements of Escherichia coli: a family of genetic composites  
 RT each encoding a large mosaic protein.";  
 RL Mol. Microbiol. 12:865-871(1994).  
 CC -1- FUNCTION: RHA ELEMENTS HAVE A NONESSENTIAL FUNCTION. THEY MAY  
 CC PLAY AN IMPORTANT ROLE IN THE NATURAL ECOLOGY OF THE CELL.  
 CC -1- DOMAIN: EACH RHA APPEARS TO CONSIST OF A HIGHLY CONSERVED 141 kDa  
 CC AMINO FRAGMENT FOLLOWED BY A HIGHLY DIVERGENT CARBOXY TERMINUS.  
 CC -1- SIMILARITY: BELONGS TO THE RHA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL, L19044; AAC95065.1; -  
 DR EMBL, U00039; AAB18570.1; -  
 DR EMBL, AE000437; AAC76617.1; -  
 DR PIR, G65159; G65159.  
 DR EcoGene; EGI0846; rhaA.  
 DR InterPro; IPR001826; RHA.  
 DR Pfam; PF03527; RHA; 1.  
 DR PRINTS; PR00394; RHA; 1.  
 DR TRIPFAMS; TIGR01643; YD repeat 2x; 15.  
 KY Signal; Multigene family; Transmembrane; Repeat; Complete proteome.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1377  
 FT TRANSMEM 28 55  
 FT DOMAIN 330 1186 28 X APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 330 352 1.  
 FT REPEAT 353 374 2.  
 FT REPEAT 375 417 3.  
 FT REPEAT 418 438 4.  
 FT REPEAT 439 460 5.  
 FT REPEAT 461 481 6.  
 FT REPEAT 482 502 7.  
 FT REPEAT 503 525 8.  
 FT REPEAT 526 546 9.  
 FT REPEAT 547 567 10.  
 FT REPEAT 568 588 11.  
 FT REPEAT 589 609 12.  
 FT REPEAT 610 629 13.  
 FT REPEAT 630 650 14.  
 FT REPEAT 651 671 15.

FT REPEAT 672 691 16.  
 FT REPEAT 692 711 17.  
 FT REPEAT 712 734 18.  
 FT REPEAT 735 758 19.  
 FT REPEAT 808 828 20.  
 FT REPEAT 829 850 21.  
 FT REPEAT 851 871 22.  
 FT REPEAT 872 894 23.  
 FT REPEAT 895 930 24.  
 FT REPEAT 931 959 25.  
 FT REPEAT 960 984 26.  
 FT REPEAT 985 1019 27.  
 FT REPEAT 1162 1186 28.  
 SQ SEQUENCE 1377 AA; 156320 MW; 21ACA989E74200FE CRC64;  
 Query Match 2.8%; Score 109.5; DB 1; Length 1377;  
 Best Local Similarity 20.8%; Pred. No. 8.4;  
 Matches 128; Conservative 58; Mismatches 209; Indels 219; Gaps 35;  
 QY 57 AAKPGDITSLGGG---GAARVVRSGTGYAA---AGL--HRYLDPGCGHYAMGSQL 108  
 DB 6 AARQG-DMTQYGGSIYQSGAGVRIIGAPTVAGSCVCGVSGHPVNPILGAKYLPGETDI 64  
 QY 109 RLPRPLPAV-----PG-----ELTEAT-----PNRYRYQ 133  
 DB 65 ALPGPLPILSRYSYRTTPAPVGLSGWGMKPADILQLRDNLLISDNGRSLYFE 124  
 QY 134 NVC--TQSYFVWDMARWEREIDMALNGINIALAMSGOEAIQSV-----YLA 181  
 DB 125 HLPFGEDGVS---RSESLMVRGGVALDGBRHIALMQALPELRSLPRRYA 175  
 QY 182 LGTQAEINEFTPGAPFLAMGRMNLHTW-----DGLPPSMWIKQLYGHRV 230  
 DB 176 TNSPQ-----GP-----MWLIG-----MCRVPEADDEVLPALPP-----YRVL 209  
 QY 231 DQW-RSFGMTPLVPAPAGHVEAVTVFPQVNTKXSGWGHFNCYSYCSFLAPEDPIFP 289  
 DB 210 TGVDFRFRGRTQTHR-----EAGFGSGEITGVTDGAWHFRLVLTQORALE----- 258  
 QY 290 IIGSLFRELKRFQGDHITYGADTFNEMQPSSEPEYLAATVAEAMTAUVEAWTVLL 349  
 DB 259 -----ARQALSGGTE-----PSAFPTLLGQYTE--YGRDNGILSLAWL-- 296  
 QY 350 QGWLFOHOPQFMGPAPIRAV---LGAVPRGLVLIDLPASQPVYTRTASFGQCFIWC 405  
 DB 297 -----THDPEY--PENLPAPALIVRYGMPTRGELAV-----VYDRSGK-- 311  
 QY 406 MANNFGGNHGLFALRA--VNGCEPARLFPNSITVGTGMAPBGIS----- 449  
 DB 332 QVRSFTYDDKYRGMVAHHRHTGRFEIRYRYSDDGRTEQLNPAGLSYTYQEKDRIITTD 391  
 QY 450 ---ONEVVSVMALGMRK-----DVPDLAAWVSFAARRYGVSHPDA 490  
 DB 392 SLRREVLRH-QGBAGLKVYKKEHADSGVTQSGPFAVGRIRA--QTDAAGRITTESPDV 448  
 QY 491 GAAMRLILSVNCSGEC---RGHNSPLVRRP-SLOMNTSITWNSRDVPEAMFLILIT 545  
 DB 449 VTG---LITRITTPDRAAFAFYNNHNOULTSATGPDLTEL-----REYDELGLRIG 498  
 QY 546 SAPSLATSPAFRYD 559  
 DB 499 TAPD-GDITRYRYD 511  
 RESULT 7  
 ID RHB ECOLI STANDARD; PRT; 1411 AA.  
 AC P16917; P76701;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE RHB protein precursor.  
 GN RHB OR B3482.

OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxId=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=93259920; PubMed=8387990;  
 RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;  
 RT "Rhs elements of Escherichia coli K-12: complex composites of shared  
 RT and unique components that have different evolutionary histories.";  
 RI J. Bacteriol. 175:2799-2808(1993).  
 RN [2]  
 RP REVISION TO 405.  
 RA Hill C.W.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94316500; PubMed=8041620;  
 RA Sofia H.V., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
 RT region from 76.0 to 81.5 minutes.";  
 RI Nucleic Acids Res. 22:2576-2586(1994).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RI Science 277:1453-1474(1997).  
 RN [5]  
 RP SEQUENCE OF 1-100 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=89123133; PubMed=2644231;  
 RA Sadosky A.B., Davidson A., Lin R.J., Hill C.W.;  
 RT "rhs gene family of Escherichia coli K-12.";  
 RI J. Bacteriol. 171:636-642(1989).  
 RN [6]  
 RP SEQUENCE OF 1221-1411 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90094253; PubMed=2403547;  
 RA Feulner G., Gray J.A., Kirschman J.A., Lehner A.F., Sadosky A.B.,  
 RA Vlazny D.A., Zhang J., Zhao S., Hill C.W.;  
 RT "Structure of the rhs locus from Escherichia coli K-12 and  
 RT comparison of rhsA with other members of the rhs multigene family.";  
 RI J. Bacteriol. 172:446-456(1990).  
 RN [7]  
 RP REVIEW.  
 RP MEDLINE=95020608; PubMed=7934896;  
 RA Hill C.W., Sandt C.H., Vlazny D.A.;  
 RT "Rhs elements of Escherichia coli: a family of genetic composites  
 RT each encoding a large mosaic protein.";  
 RI Mol. Microbiol. 12:865-871(1994).  
 RN [8]  
 RP FUNCTION: Rhs ELEMENTS HAVE A NONESSENTIAL FUNCTION. THEY MAY  
 RP PLAY AN IMPORTANT ROLE IN THE NATURAL ECOLOGY OF THE CELL.  
 RP -1- DOMAIN: EACH Rhs APPEARS TO CONSIST OF A HIGHLY CONSERVED 141 KDA  
 RP AMINO FRAGMENT FOLLOWED BY A HIGHLY DIVERGENT CARBOXY TERMINUS.  
 RP -1- SIMILARITY: BELONGS TO THE Rhs FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L02370; AAC61883.1; -  
 CC EMBL; U00039; AAB18457.1; -

DR EMBL; AB000424; AAC76507.1; -.  
 DR PIR; E65145; E65145.  
 DR Ecocore; EG10847; rhsB.  
 DR InterPro; IPR001826; rhs.  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF03527; rhs; 1.  
 DR PRINTS; PR00394; RhsPROTEIN.  
 DR TIGRFAMs; TIGR01643; YD\_repeat\_2x; 15.  
 KW Signal; Multigene family; Transmembrane; Repeat; Complete proteome.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1411  
 FT TRANSMEM 28 55  
 FT DOMAIN 330 1186  
 FT REPEAT 330 352  
 FT REPEAT 353 374  
 FT REPEAT 375 417  
 FT REPEAT 418 438  
 FT REPEAT 439 460  
 FT REPEAT 461 481  
 FT REPEAT 482 502  
 FT REPEAT 503 525  
 FT REPEAT 526 546  
 FT REPEAT 547 567  
 FT REPEAT 568 588  
 FT REPEAT 589 609  
 FT REPEAT 610 629  
 FT REPEAT 630 650  
 FT REPEAT 651 671  
 FT REPEAT 672 691  
 FT REPEAT 692 711  
 FT REPEAT 712 734  
 FT REPEAT 735 758  
 FT REPEAT 808 828  
 FT REPEAT 829 850  
 FT REPEAT 851 871  
 FT REPEAT 872 894  
 FT REPEAT 895 930  
 FT REPEAT 931 959  
 FT REPEAT 960 984  
 FT REPEAT 985 1019  
 FT REPEAT 1162 1186  
 FT CONFLICT 1130 1130 Q->K (IN REF. 4).  
 SQ SEQUENCE 1411 AA; 159394 MM; 567AC4EB713D9E07 CRC64;  
 Query Match 2.8%; Score 109.5; DB 1; Length 1411;  
 Best Local Similarity 20.8%; Pred. No. 8.6;  
 Matches 128; Conservative 58; Mismatches 209; Indels 219; Gaps 35;  
 QY 57 AKKPGIDYISLGGG---GAARVVRGSGVAA--AGL--HRYLRDSCGHVAMSGSL 108  
 DB 6 AARQG-DMTQYGGSTVQSGAGVRIGAPTVGACVCGGVTSGHPVNPILGAKVLPGETDI 64  
 QY 109 RLPRPLPAV-----PG-----ELTEAT-----PNEYRYQ 133  
 DB 65 ALPGPLPFLSTSYRTKTPAPVGLSPGKPKPADIRLQLRDNTLLISDNGSLTYFE 124  
 QY 134 NYC--TQSYFVFWDMANREITDMALNGINLALANGQALNQRV-----YLA 181  
 DB 125 HLPFSDGYS-----RSSSLVLRGVGAKLDEGRLAALWQALPEELRSLPHRYLA 175  
 QY 182 LGLTQALINEFTGPAFLAMGMLHTM-----DGLPSPWHIKOLYLQHRVL 230  
 DB 176 TNSPQ-----GF-----WMLG-----WCEVPPEADVLPAPLP-----YVL 209  
 QY 231 DQM-RSFGKTPVLPAPAGHPVAVTRVPQVNVTKMGSGWGFNCSYSGSFLAPDPFP 289  
 DB 210 TGLVVRFRGTQTFHR-----EAAQGFSGEITGVTDGMRHRRLVLTQQAPAE----- 258  
 QY 290 IIGSLFLRELKEFGDHIYGADTFENMQPSSSEPSYLAATYVEAMTAVDEAVWL 349  
 DB 259 -----ARQQAISGTE-----BSAFPDTLPQTE--YGRDNGRLSAVWL- 296  
 QY 350 QGMLFQHQPFQWGAQIRAV-----LGAVPRGRLLVLDLFASSQPFYTRTASFQGPFTWC 405

```

Db 297 ----THDPEY--PENI:PAAPLVYVGTWPTGELAV-----VYDRSGK----- 331
Qy 406 MAFNFGNGHGLFGALFA--VNGGPEAARLFNFTWGTGNAPEGIS----- 449
Db 332 QVRFSTYDDKRGKGMVARRHGRPEIRYRSDGRVTEQLNPAQLSTTYQEKDRITTD 331
Qy 450 ---QNEVYYSIMAEIGWRK-----DPVPLAAWVTSFAARYGVSHDA 490
Db 392 SLDRREVLAHT-QGEGAGLGRVYKKEHADGVSYSQFDVAGRLFA--CTDAAGRTTESPDV 448
Qy 491 GAARWLLIRSVYNSGEGAC-----RGHNSPLVRRP-SLQWMTSTWNRSDVFEAMRLILT 545
Db 449 VTG--LITRTITPDGRASAFYVYHNNQTSATGPDGLE-----RREYDELGRLIOE 498
Qy 546 SAPSLATSPAFRYD 559
Db 499 TAPD-GDITRRYRD 511

RESULT 8
ID CATA_SALTY STANDARD; PRT; 726 AA.
AC P17750;
DT 01-AUG-1990 (Rel. 15, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Peroxidase/catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)
DE (Hydroperoxidase 1).
DE KATG OR STM4106.
OS Salmomella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmomella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=91117169; PubMed=2277629;
RA Loewen P.C., Stauffer G.V.;
RT "Nucleotide sequence of katG of Salmomella typhimurium LT2 and
RT characterization of its product, hydroperoxidase I."
RL Mol. Genet. 224:147-151(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Potwolk S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmomella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
CC -1- FUNCTION: Bifunctional, exhibiting both a catalase and
CC broad-spectrum peroxidase activities.
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -1- COFACTOR: Binds 2 prohememe IX and 2 iron ions per tetramer.
CC -1- SUBUNIT: Homotrimer.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
CC -1- SIMILARITY: Contains 1 heme regulatory motif (HRM) repeat.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@sib.ch).
CC EMBL; X53001; CAA37187.1;
CC EMBL; AF008891; AAL2946.1;

```

```

DR PIR; S12039; CSEBHT.
DR HSSP; P48534; IAPX.
DR StryGene; SG10190; katG.
DR InterPro; IPR000763; Bac. class/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PRO0486; PEROXIDASE.
DR TIGRfam; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR PROSITE; PS00873; PEROXIDASE_4; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW complete proteome.
FT ACT_SITE 102 102
FT ACT_SITE 106 106
FT METAL 267 267
FT REPEAT 15 20
FT CONFIDENT 71 71
FT CONFIDENT 136 136
FT CONFIDENT 223 223
FT CONFIDENT 228 228
FT CONFIDENT 233 233
FT CONFIDENT 233 233
FT CONFIDENT 275 275
FT CONFIDENT 545 545
FT CONFIDENT 549 549
FT CONFIDENT 553 553
FT CONFIDENT 561 561
FT CONFIDENT 634 634
SQ SEQUENCE 726 AA; 79656 MW; 7C4BA4439E9FAB CRC64;

Query Match 2.7%; Score 108; DB 1; Length 726;
Best Local Similarity 17.7%; Pred. No. 4.6;
Matches 134; Conservative 73; Mismatches 196; Indels 352; Gaps 33;

Qy 86 AAGLRHYLRDPCGCHVANSQSGLRPRUPAVPGELTEATNR-----YRYNQVCTQ 138
Db 26 SAGACTASRD-----WPNQLRV-----DLNQHRSNPLGEDPYRKEFKL 69
Qy 139 SYSFV-----WMDARERELDMALNGINLALWS----- 169
Db 70 DYSALKGDIKALLTDSQW-----PADMGSYGLTIRAHMHAGTYRSDRGGAQR 122
Qy 170 -----GQEAIWQVYIALGLTQAEINEFFTPAF 198
Db 123 GQORFAPLNSWPDNVLDPKARLLWPVKOKYQKISMDLFLAGNVALENGSEFT---- 178
Qy 199 LAMGGMNLHTWDGLPSPWHIKOYLQHRVLDQWRSGMTFVLPFAFGHVEATRVFP 258
Db 179 PEGG-AGREDVPEPLDVWGDSEKALTHR-----H-PEALAKA-- 215
Qy 259 QVAVTKGSGWGHFNCSYSCSFLAPEDPIFP-IGSLFPRERIKERTDHIYGADTFNEMQ 318
Db 216 PLGATEMG-----LIYVNEGP-----DH----- 234
Qy 319 PRSSSPYLAATTVYEAATVAVTEAVYWLQGMFLPQHPQWGPAGQIRAVYGAIVPRGL 378
Db 235 --SGEPLAAMAAIRATFGMGKNDFTALLAG--GHTLKGTAALAAASHVADP--- 285
Qy 379 LVLDLFAASQPVYRTASTAFQGGPFIWCMLNFGNGHGLFGALVANGPEAA----- 430
Db 286 -----EAPF-----EAQGLGMASYSQG-----VGA-DATSLSEVWVOTPTQW 325
Qy 431 -----RLFPNSTWGTGMAPEGISQNEVYS-----LMAEIGWRK 466
Db 326 SNYFENLFKXEMV--QTRSPGALCFEAVVAPDILPPFPSPKRRKPTMTVTLTLTFD 383
Qy 467 P-----VPLDAWVTSFAARYGVSHPDAGAAWLLIRSVYNSGEGACRGNRSPL 517
Db 384 PEFEKISRFLNDPQAFNPAFAKFKLTHDMGXAYI----- 423
Qy 518 VARPSLQWNTSIWNRSDVFEAMRLILTAPSLATSPAFRYDLDL-----TROAVOEL 571
Db 424 --GPEVPEKEDLIWQD-----PLPQPLVQPT-QEDIIWLKAAIASGSISEM 467

```

```

QY 572 VSLVYEERASAYLSKELASLRLAG-----GVLAYELLPALDEVLT 610
Db 468 VSV-----AMAS---ASTFRGDDKRGANGARLAMPORDVNAVAALVPLEKIQ 517
QY 611 ASDSRFLGSG-----WLEQARAAA-----VSEAEAD 636
Db 518 KTKVKASLADIIVLAGVGIQAAAAGVSIYFAPAGVDAQDQDIEMFSLPIAD 577
QY 637 FYEQ-----NSRYOLTMGPEGNLT 656
Db 578 GFRNRYRRLDVSTTESLILDKAQQLTLPAPMTVL 612

RESULT 9
CATA_ECOLI STANDARD; PRT; 726 AA.
AC P13023; 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peroxidase/catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)
DE (Hydroperoxidase 1).
GN KATG OR B3942.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88314956; PubMed=3045098;
RA Triggs-Raine B.U., Doble B.W., Mulvey M.R., Sorby P.A., Loewen P.C.;
RT "Nucleotide sequence of katG, encoding catalase HPI of Escherichia
RT coli."
RL J. Bacteriol. 170:4415-4419 (1988).
RN [2]
RP SEQUENCE OF 1-339 FROM N.A.
RX STRAIN-K12 / MG1655.
RX MEDLINE=93347969; PubMed=8346018;
RA Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;
RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT region from 87.2 to 89.2 minutes."
RL Nucleic Acids Res. 21:3391-3398 (1993).
RN [3]
RP SEQUENCE OF 309-726 FROM N.A.
RX STRAIN-K12 / MG1655.
RX MEDLINE=94089392; PubMed=8265357;
RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.U.,
RA Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes."
RL Nucleic Acids Res. 21:5408-5417 (1993).
RN [4]
RP FUNCTION: Bifunctional, exhibiting both a catalase and
RN broad-spectrum peroxidase activities.
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -1- COFACTOR: Binds 2 prothemo IX and 2 iron ions per tetramer.
CC -1- SUBUNIT: Homotetramer.
CC -1- INDUCTION: By hydrogen peroxide.
CC -1- PMT: The N-terminus is blocked.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
CC -1- SIMILARITY: Contains 1 heme regulatory motif (HRM) repeat.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M21516; AAA24040.1; -
DR EMBL; L19201; AAB03074.1; -

```

```

DR EMBL; U00006; AAC43048.1; -
DR EMBL; AE000468; AAC76924.1; -
DR PIR; A65201; CSECHP.
DR HSSP; P00431; 1CYF.
DR SWISS-2DPAGE; P13023; COLI.
DR Ecogene; EG10511; katG.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase_1.
DR PRINTS; PRO0458; PEROXIDASE.
DR TIGRFAMs; TIGR00198; cat_per_HPI_1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR PROSITE; PS00873; PEROXIDASE_4; 1.
DR Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW Complete proteome.
FT ACT SITE 102 102 BY SIMILARITY.
FT ACT SITE 106 106 BY SIMILARITY.
FT METAL 267 267 IRON (HEME AXIAL LIGAND).
FT REPEAT 15 20 HRM.
FT CONFLICT 621 621 A -> G (IN REF. 1).
SQ SEQUENCE 726 AA; 80023 MW; 24D32EBED5D59BD6 CRC64;

Query Match 2.7%; Score 104.5; DB 1; Length 726;
Best Local Similarity 18.3%; Pred. No. 8.2;
Matches 125; Conservative 66; Mismatches 169; Indels 303; Gaps 32;

QY 86 AAGLHRYRDPGCGHVAWSGQLRPLPAVPGELTEATNR-----YR----- 130
Db 26 SAGAGTTRD-----WMPNQLRV-----DLNQHNRNPNLGEDFDYKERSKL 69
QY 131 -YY-----QNVQTSYFVWMDARWEREDMNLNGINLALMS----- 169
Db 70 DYGLKAKLKLITRSQ--WW-----PADMGSYAGLFIRMAHNGAGTYSIDGRGA 120
QY 170 -----QGEAIWQRYTALGLTQAEINEFTTGP 196
Db 121 GCGQRFAPLNSWPDNVSILDKARLLWPIKOKYQKISWALFLTAGVALENGFR-- 178
QY 197 AFLANGRGNIHTMDGCLPSPSMHIKQLYLQHRVIDQNRFSQMTPLPAFAGHVEAVTRV 256
Db 179 -FGFG-AGREDWEPDLDVWMDKAVLTR-----H-PBALAKA 215
QY 257 FQVAVVTQSGWGHNCSSCSFLLAPDPPIPTIGSLFLRELKEFGTDHIYGAOTFNE 316
Db 216 -PLGATENG-----LIYVNEGP-----DH----- 234
QY 317 NQPSSESYLAATTAYEAMTAVDTAAVWLQGLFQHPQFQWPAQIAVIGAVPRG 376
Db 235 ---GGEPLSAALAIRATFGNMGNDDETVALLIAGHTLIGTHAGPT--SNVGPDP-- 285
QY 377 RLIVLDLFAESQFVYTR-----TASFOGPFIMCMLHNGGNH--GLF-- 417
Db 286 -----EAPIEGQIGMASTYSGVGADATISLEVWVTQPTQMSVYFENLFKY 336
QY 418 -----GAL--EAVNGGPEAARLFPNSTWVGMAPEGISQNEVYYSIMAEIGWKD 466
Db 337 EMWQTRSPAGAIQFRAV-----DAPEIIPD-----PFDPSKKRKCTMLVTDLTLEFD 383
QY 467 P-----VPLDAWTSFAARRYGVSHPDAGAMWLLRSYVNCSGEACRGNRSPL 517
Db 384 DEFKISRRLNDPQAFNEAPARAFKLTTHDMGKSYI----- 423
QY 518 VERPELQWNTSIWNRSDVFEAMRLLTSAPSLATSPAFRDLDL-----TROAVQEL 571
Db 424 -GPEVPEEDILWQD-----PLQPIYVPL-EODIIDLKPAIDSGLSVSEL 467
QY 572 VSLVYEERASAYLSKELASLRLAG-----GVLAYELLPALDEVLT 610
Db 468 VSV-----AMAS---ASTFRGDDKRGANGARLAMPORDVNAVAALVPLEKIQ 517
QY 611 ASDSRFLGSGWLEQARAAVSEA 633

```

DB 518 KESKASLADITVAGVGEKA 540

RESULT 10

P85B RAT STANDARD; PRT; 722 AA.

AC 063788;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Phosphatidylinositol 3-kinase regulatory beta subunit (PI3-kinase p85-beta subunit) (Ptdins-3-kinase p85-beta).

DE PIK3R2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RP SEQUENCE FROM N.A.

RC STRAIN=Wistar; TISSUE=Brain;

RX MEDLINE=9621497; PubMed=6621382;

RA Inukai K., Anai M., Vandreva E., Hosaka T., Katagiri H., Funaki M., Fukushima Y., Ogihara T., Yazaki Y., Kikuchi M., Oka Y., Asano T.;

RT "A novel 55-kDa regulatory subunit for phosphatidylinositol 3-kinase structurally similar to p55PIK is generated by alternative splicing of the p85alpha gene.";

RL J. Biol. Chem. 271:5317-5320(1996).

CC -1- FUNCTION: BINDS TO ACTIVATED (PHOSPHORYLATED) PROTEIN-TYROSINE KINASES, THROUGH ITS SH2 DOMAIN, AND ACTS AS AN ADAPTER, MEDIATING THE ASSOCIATION OF THE P110 CATALYTIC UNIT TO THE PLASMA MEMBRANE.

CC -1- SUBUNIT: HETERODIMER OF A P110 (CATALYTIC) AND A P85 (REGULATORY) SUBUNITS.

CC -1- SIMILARITY: BELONGS TO THE PI3K P85 SUBUNIT FAMILY.

CC -1- SIMILARITY: Contains 1 Rho-GAP domain.

CC -1- SIMILARITY: Contains 2 SH2 domains.

CC -1- SIMILARITY: Contains 1 SH3 domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; D64046; BAA10926.1; -

DR HSSP; P23727; 2PNB.

DR InterPro; IPR001720; PI3K\_kinase\_P85.

DR InterPro; IPR000198; RhoGAP.

DR InterPro; IPR000980; SH2.

DR InterPro; IPR001452; SH3.

DR Pfam; PF00620; RhoGAP; 1.

DR Pfam; PF00017; SH2; 2.

DR Pfam; PF00018; SH3; 1.

DR PRINTS; PR00678; PI3KINASEP85.

DR PRINTS; PR00401; SH2DOMAIN.

DR PRODOM; PD000093; SH2; 2.

DR SMART; SM00324; RhoGAP; 1.

DR SMART; SM0252; SH2; 2.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS50238; RHO\_GAP; 1.

DR PROSITE; PS50001; SH2; 2.

DR PROSITE; PS50002; SH3; 1.

DR SH3 domain; SH2 domain; Repeat.

FT DOMAIN 4 80 SH3.

FT DOMAIN 112 289 RHO-GAP.

FT DOMAIN 324 419 SH2.1.

FT DOMAIN 616 710 SH2.2.

SO SEQUENCE 722 AA; 81328 MW; 12083688B9F670C95 CRC64;

Query Match 2.6%; Score 103; DB 1; Length 722;

Best Local Similarity 21.6%; Pred. No. 10;

Matches 87; Conservative 38; Mismatches 134; Indels 144; Gaps 19;

QY 74 RVRVNGSTGVAAAAGLHRYLDFCGCHV-----AMSGQLRPPPLPAVF----- 118

DB 63 RTRORQ-----DFPGTYVEFLGPAVALAPGPRPRGRPLPAPFLDGPSES 107

QY 119 -----GELTE-----ATENRRYRYQNVCTOSYSFVW-----DMA-----RWER 152

DB 108 GHTLASLAEQSPSPSAPPILVKLIEALQALDEFFSRPRLPAPRDMSSLDLEQMDR 167

QY 153 EIDMMALNGINIALAMSGQEAIWQRYVALGITQAEINEFF-----TGPAFLANGMGN 206

DB 168 TLLYDAVVGFFLLALP-----AAVTPEAAEEAAYANREVTGPGVGLYLE----- 210

QY 207 LHTWDGPIPPSWHINQVLYQHRVLDQMSFGMTPLPAPAGHPVAVRVFPQVAVTMKG 266

DB 211 -----PTPLPLQALTLRPLQLHGRVARRASPATAVH--ALMSAFGL----- 253

QY 267 SWGHFNCSYSCSFLIAPED-----PIFFIIGSLFRLRLKEFGTDHIYGADTNE 316

DB 254 -----LLNAPPGEGRGDSEPADPFV---LLERLVGE---HVDEQDTAPF 294

QY 317 MPPSPSESYLAATTAYTE-AMTAVDTEAVVLLQGLFQHPQPMGP---AQIAYVGA 372

DB 295 ALPP--KSKVAPAPETALANGSTSLQDAEW-----YMGDISREEVNERLUD 340

QY 373 VPRGRLLVLDLPAESQPYVTRTASFGQPFICMLNFGNGNG 415

DB 341 TPDGTFIVRDASKIKQGEYTLTRKGNKKLIKVHR-DGHYHG 382

RESULT 11

YMCA ECOLI STANDARD; PRT; 698 AA.

ID YMCA ECOLI

AC P75882;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical 1 lipoprotein ymcA precursor.

GN YMCA OR B0984.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RA "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=97061202; PubMed=8905232;

RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kaishimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N., Sampaio G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano W., Horiiuchi T.;

RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";

RL DNA Res. 3:117-155(1996).

CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor (Probable).

CC -1- SIMILARITY: STRONG, TO E. COLI YUBH.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AB000300; AAC74069.1; -  
 DR EMBL; D90735; BAA35749.1; -  
 DR EMBL; D90736; BAA36124.1; -  
 DR PIR; F64839; F64839.  
 DR EcoGene; EG13729; ymcA.  
 DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.  
 KM Hypothetical protein; Membrane; Lipoprotein; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 698 HYPOTHETICAL LIPOPROTEIN YMC.  
 FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).  
 SQ SEQUENCE 698 AA; 78687 MW; 37F71D2D57FAE87 CRC64;

Query March 2.6%; Score 101; DB 1; Length 698;  
 Best local similarity 19.2%; Pred. No. 14; Indels 300; Gaps 39;  
 Matches 146; Conservative 75; Mismatches 239;

QY 108 LRLRPLPAVVG--LLEATPNRYRYO-----NYCTOSYSPVWMDA- 148  
 DB 43 LQWNAIAPBEGSVVYRNDQYRFYSTVALFPMLEGTIRYTDVTRKYS-QMEDFSG 101  
 QY 149 -----RMREIDMAMNGLNGLAMSGEALWCVYALGL-----T 185  
 DB 102 DQSYKDSFDFKRLMEEGY-----WLPVAGKDKDAGTGFGDEYVASK 148  
 QY 186 QAEINEFFTPGPAFLAMRGKGLHTWDGRLPSPWHIKOLYLOHNV-----LDQMS 235  
 DB 149 QAGSFDFTLG--MAMVAGNAGNITMP--CRVSKVC-HAASHDAGDISFSDFIFG 201  
 QY 236 ---FG--MTPVLP--FAHVEAVTRVFPVONTTKMGSGWGHNC 273  
 DB 202 PASIFGIEYQTPMNPRLKLEYDGNVYQNDFAKGLPQA--SHFNVAAYVRAASMDLNL 259  
 QY 274 SYSQSFLLAPEDPFPITIGSLFRELKERTDHYV--ADTFENMOP--PSSPSYL 327  
 DB 260 SY-----ERGNLMEFRTLRTPNDLRPLRTPTKPAVQ 293  
 QY 328 AAAT--TAYEANTAV-----DTEAVMLQGLFQHOPQFQSPACIRAVLGAVR 375  
 DB 294 PAFSEGLQYTTVAQNTALKYNAQFAPRIQLRDKTLVMSGOQY--KYRDSRAVDNR 349  
 QY 376 -GRLLVLDLPAESQPVYTRTASFGQGFIMQMLNFGNNGLFGALEAVNGGEARLRF 434  
 DB 350 ANRLVNNL--P-----QGVETI--STQKEHMAWT 378  
 QY 435 NSTWVGT--GMAPEGISQNEVVYSLMAELGRKDPVDDLAAMVTSFAARVGVSHP 488  
 DB 379 TETIVASLRKQLATAP--QOSEPLQOQVVE-----AEDLSAFGRGRIHEDPFSYS 428  
 QY 489 DAGAAMRLLRSTVNGSGEACGRHNSPLVRRPDLQNTSIW-----YNR 533  
 DB 429 -----NPFLLSQSLGPEDEPFYFQGLMSARVFTDHLLDGIFETNIYNN 475  
 QY 534 SDVEAWRLILTSAPSLATSPAFYDDLDLRQAV-----QELVSLYTEARSAY 583  
 DB 476 YDKKSSIL--PASTILPRVYTHRIDVYNDVYNNIQAAYFADLNGSFGQVYGGY 530  
 QY 584 LSKELASLRAGVLAELLALDEVIASDSRFLGSMLEQAPRAAVSEADPFYEQNR 643  
 DB 531 LETWYAGV--GSEILVRLP-----DACWMLG-----VDVYVVCQRD- 564  
 QY 644 YQLTLMGEGNILLYANKQLAGLVANYVTPRWRFLFLALVDVSAQGI PPOHQFQKVVQ 703  
 DB 565 -----WNNMAFTYISPD--TGFTATVNP-----PILNGVIMK----- 596  
 QY 704 LEQAFVLSKORYPSQPGDVTDLAKKIFLKYYPGWVAGSW 743  
 DB 597 -----LSVGOYLAKDKGATIDVAK--RDSGVAVGWSV 626

RESULT 12  
 ID\_056 STRAT STANDARD; PRT; 3519 AA.  
 AC 007017;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Oleandomycin polyketide synthase, modules 5 and 6.  
 GN ORF5.  
 OS Streptomyces antibioticus.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 OK NCBI\_taxid=1890;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150470; PubMed=8107683;  
 RA Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;  
 RT "Characterisation of a Streptomyces antibioticus gene encoding a type  
 RT I polyketide synthase which has an unusual coding sequence."  
 RL Mol. Gen. Genet. 242:358-362 (1994).  
 CC -1- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN  
 CC LACTONE RING.  
 CC -1- COPOLYMER: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.  
 CC -1- SIMILARITY: Contains 2 acyl carrier domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L09654; AAA16695.1; -  
 DR PIR; S43048; S43048.  
 DR HSSP; P25715; 1MLA.  
 DR InterPro; IPR001227; Ac transferase.  
 DR InterPro; IPR000794; Ketoacyl-synt.  
 DR InterPro; IPR006163; Pp-bind.  
 DR InterPro; IPR006162; Ppant-attach.  
 DR InterPro; IPR001031; Thioesterase.  
 DR Pfam; PF00698; Acyl\_transf. 2.  
 DR Pfam; PF00109; ketoacyl-synt; 2.  
 DR Pfam; PF02801; ketoacyl-synt; C; 2.  
 DR Pfam; PF00550; pp-binding; 2.  
 DR Pfam; PF00875; Thioesterase; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.  
 DR PROSITE; PS00606; B\_KETOACYL SYNTHASE; 2.  
 DR PROSITE; PS00075; ACP DOMAIN; 2.  
 DR Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP;  
 KW Phosphopantetheine; Multifunctional enzyme; Repeat.  
 FT DOMAIN 1  
 FT DOMAIN 2  
 FT DOMAIN 3  
 FT DOMAIN 4  
 FT DOMAIN 5  
 FT DOMAIN 6  
 FT DOMAIN 7  
 FT DOMAIN 8  
 FT DOMAIN 9  
 FT DOMAIN 10  
 FT DOMAIN 11  
 FT DOMAIN 12  
 FT DOMAIN 13  
 FT DOMAIN 14  
 FT DOMAIN 15  
 FT DOMAIN 16  
 FT DOMAIN 17  
 FT DOMAIN 18  
 FT DOMAIN 19  
 FT DOMAIN 20  
 FT DOMAIN 21  
 FT DOMAIN 22  
 FT DOMAIN 23  
 FT DOMAIN 24  
 FT DOMAIN 25  
 FT DOMAIN 26  
 FT DOMAIN 27  
 FT DOMAIN 28  
 FT DOMAIN 29  
 FT DOMAIN 30  
 FT DOMAIN 31  
 FT DOMAIN 32  
 FT DOMAIN 33  
 FT DOMAIN 34  
 FT DOMAIN 35  
 FT DOMAIN 36  
 FT DOMAIN 37  
 FT DOMAIN 38  
 FT DOMAIN 39  
 FT DOMAIN 40  
 FT DOMAIN 41  
 FT DOMAIN 42  
 FT DOMAIN 43  
 FT DOMAIN 44  
 FT DOMAIN 45  
 FT DOMAIN 46  
 FT DOMAIN 47  
 FT DOMAIN 48  
 FT DOMAIN 49  
 FT DOMAIN 50  
 FT DOMAIN 51  
 FT DOMAIN 52  
 FT DOMAIN 53  
 FT DOMAIN 54  
 FT DOMAIN 55  
 FT DOMAIN 56  
 FT DOMAIN 57  
 FT DOMAIN 58  
 FT DOMAIN 59  
 FT DOMAIN 60  
 FT DOMAIN 61  
 FT DOMAIN 62  
 FT DOMAIN 63  
 FT DOMAIN 64  
 FT DOMAIN 65  
 FT DOMAIN 66  
 FT DOMAIN 67  
 FT DOMAIN 68  
 FT DOMAIN 69  
 FT DOMAIN 70  
 FT DOMAIN 71  
 FT DOMAIN 72  
 FT DOMAIN 73  
 FT DOMAIN 74  
 FT DOMAIN 75  
 FT DOMAIN 76  
 FT DOMAIN 77  
 FT DOMAIN 78  
 FT DOMAIN 79  
 FT DOMAIN 80  
 FT DOMAIN 81  
 FT DOMAIN 82  
 FT DOMAIN 83  
 FT DOMAIN 84  
 FT DOMAIN 85  
 FT DOMAIN 86  
 FT DOMAIN 87  
 FT DOMAIN 88  
 FT DOMAIN 89  
 FT DOMAIN 90  
 FT DOMAIN 91  
 FT DOMAIN 92  
 FT DOMAIN 93  
 FT DOMAIN 94  
 FT DOMAIN 95  
 FT DOMAIN 96  
 FT DOMAIN 97  
 FT DOMAIN 98  
 FT DOMAIN 99  
 FT DOMAIN 100  
 FT DOMAIN 101  
 FT DOMAIN 102  
 FT DOMAIN 103  
 FT DOMAIN 104  
 FT DOMAIN 105  
 FT DOMAIN 106  
 FT DOMAIN 107  
 FT DOMAIN 108  
 FT DOMAIN 109  
 FT DOMAIN 110  
 FT DOMAIN 111  
 FT DOMAIN 112  
 FT DOMAIN 113  
 FT DOMAIN 114  
 FT DOMAIN 115  
 FT DOMAIN 116  
 FT DOMAIN 117  
 FT DOMAIN 118  
 FT DOMAIN 119  
 FT DOMAIN 120  
 FT DOMAIN 121  
 FT DOMAIN 122  
 FT DOMAIN 123  
 FT DOMAIN 124  
 FT DOMAIN 125  
 FT DOMAIN 126  
 FT DOMAIN 127  
 FT DOMAIN 128  
 FT DOMAIN 129  
 FT DOMAIN 130  
 FT DOMAIN 131  
 FT DOMAIN 132  
 FT DOMAIN 133  
 FT DOMAIN 134  
 FT DOMAIN 135  
 FT DOMAIN 136  
 FT DOMAIN 137  
 FT DOMAIN 138  
 FT DOMAIN 139  
 FT DOMAIN 140  
 FT DOMAIN 141  
 FT DOMAIN 142  
 FT DOMAIN 143  
 FT DOMAIN 144  
 FT DOMAIN 145  
 FT DOMAIN 146  
 FT DOMAIN 147  
 FT DOMAIN 148  
 FT DOMAIN 149  
 FT DOMAIN 150  
 FT DOMAIN 151  
 FT DOMAIN 152  
 FT DOMAIN 153  
 FT DOMAIN 154  
 FT DOMAIN 155  
 FT DOMAIN 156  
 FT DOMAIN 157  
 FT DOMAIN 158  
 FT DOMAIN 159  
 FT DOMAIN 160  
 FT DOMAIN 161  
 FT DOMAIN 162  
 FT DOMAIN 163  
 FT DOMAIN 164  
 FT DOMAIN 165  
 FT DOMAIN 166  
 FT DOMAIN 167  
 FT DOMAIN 168  
 FT DOMAIN 169  
 FT DOMAIN 170  
 FT DOMAIN 171  
 FT DOMAIN 172  
 FT DOMAIN 173  
 FT DOMAIN 174  
 FT DOMAIN 175  
 FT DOMAIN 176  
 FT DOMAIN 177  
 FT DOMAIN 178  
 FT DOMAIN 179  
 FT DOMAIN 180  
 FT DOMAIN 181  
 FT DOMAIN 182  
 FT DOMAIN 183  
 FT DOMAIN 184  
 FT DOMAIN 185  
 FT DOMAIN 186  
 FT DOMAIN 187  
 FT DOMAIN 188  
 FT DOMAIN 189  
 FT DOMAIN 190  
 FT DOMAIN 191  
 FT DOMAIN 192  
 FT DOMAIN 193  
 FT DOMAIN 194  
 FT DOMAIN 195  
 FT DOMAIN 196  
 FT DOMAIN 197  
 FT DOMAIN 198  
 FT DOMAIN 199  
 FT DOMAIN 200  
 FT DOMAIN 201  
 FT DOMAIN 202  
 FT DOMAIN 203  
 FT DOMAIN 204  
 FT DOMAIN 205  
 FT DOMAIN 206  
 FT DOMAIN 207  
 FT DOMAIN 208  
 FT DOMAIN 209  
 FT DOMAIN 210  
 FT DOMAIN 211  
 FT DOMAIN 212  
 FT DOMAIN 213  
 FT DOMAIN 214  
 FT DOMAIN 215  
 FT DOMAIN 216  
 FT DOMAIN 217  
 FT DOMAIN 218  
 FT DOMAIN 219  
 FT DOMAIN 220  
 FT DOMAIN 221  
 FT DOMAIN 222  
 FT DOMAIN 223  
 FT DOMAIN 224  
 FT DOMAIN 225  
 FT DOMAIN 226  
 FT DOMAIN 227  
 FT DOMAIN 228  
 FT DOMAIN 229  
 FT DOMAIN 230  
 FT DOMAIN 231  
 FT DOMAIN 232  
 FT DOMAIN 233  
 FT DOMAIN 234  
 FT DOMAIN 235  
 FT DOMAIN 236  
 FT DOMAIN 237  
 FT DOMAIN 238  
 FT DOMAIN 239  
 FT DOMAIN 240  
 FT DOMAIN 241  
 FT DOMAIN 242  
 FT DOMAIN 243  
 FT DOMAIN 244  
 FT DOMAIN 245  
 FT DOMAIN 246  
 FT DOMAIN 247  
 FT DOMAIN 248  
 FT DOMAIN 249  
 FT DOMAIN 250  
 FT DOMAIN 251  
 FT DOMAIN 252  
 FT DOMAIN 253  
 FT DOMAIN 254  
 FT DOMAIN 255  
 FT DOMAIN 256  
 FT DOMAIN 257  
 FT DOMAIN 258  
 FT DOMAIN 259  
 FT DOMAIN 260  
 FT DOMAIN 261  
 FT DOMAIN 262  
 FT DOMAIN 263  
 FT DOMAIN 264  
 FT DOMAIN 265  
 FT DOMAIN 266  
 FT DOMAIN 267  
 FT DOMAIN 268  
 FT DOMAIN 269  
 FT DOMAIN 270  
 FT DOMAIN 271  
 FT DOMAIN 272  
 FT DOMAIN 273  
 FT DOMAIN 274  
 FT DOMAIN 275  
 FT DOMAIN 276  
 FT DOMAIN 277  
 FT DOMAIN 278  
 FT DOMAIN 279  
 FT DOMAIN 280  
 FT DOMAIN 281  
 FT DOMAIN 282  
 FT DOMAIN 283  
 FT DOMAIN 284  
 FT DOMAIN 285  
 FT DOMAIN 286  
 FT DOMAIN 287  
 FT DOMAIN 288  
 FT DOMAIN 289  
 FT DOMAIN 290  
 FT DOMAIN 291  
 FT DOMAIN 292  
 FT DOMAIN 293  
 FT DOMAIN 294  
 FT DOMAIN 295  
 FT DOMAIN 296  
 FT DOMAIN 297  
 FT DOMAIN 298  
 FT DOMAIN 299  
 FT DOMAIN 300  
 FT DOMAIN 301  
 FT DOMAIN 302  
 FT DOMAIN 303  
 FT DOMAIN 304  
 FT DOMAIN 305  
 FT DOMAIN 306  
 FT DOMAIN 307  
 FT DOMAIN 308  
 FT DOMAIN 309  
 FT DOMAIN 310  
 FT DOMAIN 311  
 FT DOMAIN 312  
 FT DOMAIN 313  
 FT DOMAIN 314  
 FT DOMAIN 315  
 FT DOMAIN 316  
 FT DOMAIN 317  
 FT DOMAIN 318  
 FT DOMAIN 319  
 FT DOMAIN 320  
 FT DOMAIN 321  
 FT DOMAIN 322  
 FT DOMAIN 323  
 FT DOMAIN 324  
 FT DOMAIN 325  
 FT DOMAIN 326  
 FT DOMAIN 327  
 FT DOMAIN 328  
 FT DOMAIN 329  
 FT DOMAIN 330  
 FT DOMAIN 331  
 FT DOMAIN 332  
 FT DOMAIN 333  
 FT DOMAIN 334  
 FT DOMAIN 335  
 FT DOMAIN 336  
 FT DOMAIN 337  
 FT DOMAIN 338  
 FT DOMAIN 339  
 FT DOMAIN 340  
 FT DOMAIN 341  
 FT DOMAIN 342  
 FT DOMAIN 343  
 FT DOMAIN 344  
 FT DOMAIN 345  
 FT DOMAIN 346  
 FT DOMAIN 347  
 FT DOMAIN 348  
 FT DOMAIN 349  
 FT DOMAIN 350  
 FT DOMAIN 351  
 FT DOMAIN 352  
 FT DOMAIN 353  
 FT DOMAIN 354  
 FT DOMAIN 355  
 FT DOMAIN 356  
 FT DOMAIN 357  
 FT DOMAIN 358  
 FT DOMAIN 359  
 FT DOMAIN 360  
 FT DOMAIN 361  
 FT DOMAIN 362  
 FT DOMAIN 363  
 FT DOMAIN 364  
 FT DOMAIN 365  
 FT DOMAIN 366  
 FT DOMAIN 367  
 FT DOMAIN 368  
 FT DOMAIN 369  
 FT DOMAIN 370  
 FT DOMAIN 371  
 FT DOMAIN 372  
 FT DOMAIN 373  
 FT DOMAIN 374  
 FT DOMAIN 375  
 FT DOMAIN 376  
 FT DOMAIN 377  
 FT DOMAIN 378  
 FT DOMAIN 379  
 FT DOMAIN 380  
 FT DOMAIN 381  
 FT DOMAIN 382  
 FT DOMAIN 383  
 FT DOMAIN 384  
 FT DOMAIN 385  
 FT DOMAIN 386  
 FT DOMAIN 387  
 FT DOMAIN 388  
 FT DOMAIN 389  
 FT DOMAIN 390  
 FT DOMAIN 391  
 FT DOMAIN 392  
 FT DOMAIN 393  
 FT DOMAIN 394  
 FT DOMAIN 395  
 FT DOMAIN 396  
 FT DOMAIN 397  
 FT DOMAIN 398  
 FT DOMAIN 399  
 FT DOMAIN 400  
 FT DOMAIN 401  
 FT DOMAIN 402  
 FT DOMAIN 403  
 FT DOMAIN 404  
 FT DOMAIN 405  
 FT DOMAIN 406  
 FT DOMAIN 407  
 FT DOMAIN 408  
 FT DOMAIN 409  
 FT DOMAIN 410  
 FT DOMAIN 411  
 FT DOMAIN 412  
 FT DOMAIN 413  
 FT DOMAIN 414  
 FT DOMAIN 415  
 FT DOMAIN 416  
 FT DOMAIN 417  
 FT DOMAIN 418  
 FT DOMAIN 419  
 FT DOMAIN 420  
 FT DOMAIN 421  
 FT DOMAIN 422  
 FT DOMAIN 423  
 FT DOMAIN 424  
 FT DOMAIN 425  
 FT DOMAIN 426  
 FT DOMAIN 427  
 FT DOMAIN 428  
 FT DOMAIN 429  
 FT DOMAIN 430  
 FT DOMAIN 431  
 FT DOMAIN 432  
 FT DOMAIN 433  
 FT DOMAIN 434  
 FT DOMAIN 435  
 FT DOMAIN 436  
 FT DOMAIN 437  
 FT DOMAIN 438  
 FT DOMAIN 439  
 FT DOMAIN 440  
 FT DOMAIN 441  
 FT DOMAIN 442  
 FT DOMAIN 443  
 FT DOMAIN 444  
 FT DOMAIN 445  
 FT DOMAIN 446  
 FT DOMAIN 447  
 FT DOMAIN 448  
 FT DOMAIN 449  
 FT DOMAIN 450  
 FT DOMAIN 451  
 FT DOMAIN 452  
 FT DOMAIN 453  
 FT DOMAIN 454  
 FT DOMAIN 455  
 FT DOMAIN 456  
 FT DOMAIN 457  
 FT DOMAIN 458  
 FT DOMAIN 459  
 FT DOMAIN 460  
 FT DOMAIN 461  
 FT DOMAIN 462  
 FT DOMAIN 463  
 FT DOMAIN 464  
 FT DOMAIN 465  
 FT DOMAIN 466  
 FT DOMAIN 467  
 FT DOMAIN 468  
 FT DOMAIN 469  
 FT DOMAIN 470  
 FT DOMAIN 471  
 FT DOMAIN 472  
 FT DOMAIN 473  
 FT DOMAIN 474  
 FT DOMAIN 475  
 FT DOMAIN 476  
 FT DOMAIN 477  
 FT DOMAIN 478  
 FT DOMAIN 479  
 FT DOMAIN 480  
 FT DOMAIN 481  
 FT DOMAIN 482  
 FT DOMAIN 483  
 FT DOMAIN 484  
 FT DOMAIN 485  
 FT DOMAIN 486  
 FT DOMAIN 487  
 FT DOMAIN 488  
 FT DOMAIN 489  
 FT DOMAIN 490  
 FT DOMAIN 491  
 FT DOMAIN 492  
 FT DOMAIN 493  
 FT DOMAIN 494  
 FT DOMAIN 495  
 FT DOMAIN 496  
 FT DOMAIN 497  
 FT DOMAIN 498  
 FT DOMAIN 499  
 FT DOMAIN 500  
 FT DOMAIN 501  
 FT DOMAIN 502  
 FT DOMAIN 503  
 FT DOMAIN 504  
 FT DOMAIN 505  
 FT DOMAIN 506  
 FT DOMAIN 507  
 FT DOMAIN 508  
 FT DOMAIN 509  
 FT DOMAIN 510  
 FT DOMAIN 511  
 FT DOMAIN 512  
 FT DOMAIN 513  
 FT DOMAIN 514  
 FT DOMAIN 515  
 FT DOMAIN 516  
 FT DOMAIN 517  
 FT DOMAIN 518  
 FT DOMAIN 519  
 FT DOMAIN 520  
 FT DOMAIN 521  
 FT DOMAIN 522  
 FT DOMAIN 523  
 FT DOMAIN 524  
 FT DOMAIN 525  
 FT DOMAIN 526  
 FT DOMAIN 527  
 FT DOMAIN 528  
 FT DOMAIN 529  
 FT DOMAIN 530  
 FT DOMAIN 531  
 FT DOMAIN 532  
 FT DOMAIN 533  
 FT DOMAIN 534  
 FT DOMAIN 535  
 FT DOMAIN 536  
 FT DOMAIN 537  
 FT DOMAIN 538  
 FT DOMAIN 539  
 FT DOMAIN 540  
 FT DOMAIN 541  
 FT DOMAIN 542  
 FT DOMAIN 543  
 FT DOMAIN 544  
 FT DOMAIN 545  
 FT DOMAIN 546  
 FT DOMAIN 547  
 FT DOMAIN 548  
 FT DOMAIN 549  
 FT DOMAIN 550  
 FT DOMAIN 551  
 FT DOMAIN 552  
 FT DOMAIN 553  
 FT DOMAIN 554  
 FT DOMAIN 555  
 FT DOMAIN 556  
 FT DOMAIN 557  
 FT DOMAIN 558  
 FT DOMAIN 559  
 FT DOMAIN 560  
 FT DOMAIN 561  
 FT DOMAIN 562  
 FT DOMAIN 563  
 FT DOMAIN 564  
 FT DOMAIN 565  
 FT DOMAIN 566  
 FT DOMAIN 567  
 FT DOMAIN 568  
 FT DOMAIN 569  
 FT DOMAIN 570  
 FT DOMAIN 571  
 FT DOMAIN 572  
 FT DOMAIN 573  
 FT DOMAIN 574  
 FT DOMAIN 575  
 FT DOMAIN 576  
 FT DOMAIN 577  
 FT DOMAIN 578  
 FT DOMAIN 579  
 FT DOMAIN 580  
 FT DOMAIN 581  
 FT DOMAIN 582  
 FT DOMAIN 583  
 FT DOMAIN 584  
 FT DOMAIN 585  
 FT DOMAIN 586  
 FT DOMAIN 587  
 FT DOMAIN 588  
 FT DOMAIN 589  
 FT DOMAIN 590  
 FT DOMAIN 591  
 FT DOMAIN 592  
 FT DOMAIN 593  
 FT DOMAIN 594  
 FT DOMAIN 595  
 FT DOMAIN 596  
 FT DOMAIN 597  
 FT DOMAIN 598  
 FT DOMAIN 599  
 FT DOMAIN 600  
 FT DOMAIN 601  
 FT DOMAIN 602  
 FT DOMAIN 603  
 FT DOMAIN 604  
 FT DOMAIN 605  
 FT DOMAIN 606  
 FT DOMAIN 607  
 FT DOMAIN 608  
 FT DOMAIN 609  
 FT DOMAIN 610  
 FT DOMAIN 611  
 FT DOMAIN 612  
 FT DOMAIN 613  
 FT DOMAIN 614  
 FT DOMAIN 615  
 FT DOMAIN 616  
 FT DOMAIN 617  
 FT DOMAIN 618  
 FT DOMAIN 619  
 FT DOMAIN 620  
 FT DOMAIN 621  
 FT DOMAIN 622  
 FT DOMAIN 623  
 FT DOMAIN 624  
 FT DOMAIN 625  
 FT DOMAIN 626  
 FT DOMAIN 627  
 FT DOMAIN 628  
 FT DOMAIN 629  
 FT DOMAIN 630  
 FT DOMAIN 631  
 FT DOMAIN 632  
 FT DOMAIN 633  
 FT DOMAIN 634  
 FT DOMAIN 635  
 FT DOMAIN 636  
 FT DOMAIN 637  
 FT DOMAIN 638  
 FT DOMAIN 639  
 FT DOMAIN 640  
 FT DOMAIN 641  
 FT DOMAIN 642  
 FT DOMAIN 643  
 FT DOMAIN 644  
 FT DOMAIN 645  
 FT DOMAIN 646  
 FT DOMAIN 647  
 FT DOMAIN 648  
 FT DOMAIN 649  
 FT DOMAIN 650  
 FT DOMAIN 651  
 FT DOMAIN 652  
 FT DOMAIN 653  
 FT DOMAIN 654  
 FT DOMAIN 655  
 FT DOMAIN 656  
 FT DOMAIN 657  
 FT DOMAIN 658  
 FT DOMAIN 659  
 FT DOMAIN 660  
 FT DOMAIN 661  
 FT DOMAIN 662  
 FT DOMAIN 663  
 FT DOMAIN 664  
 FT DOMAIN 665  
 FT DOMAIN 666  
 FT DOMAIN 667  
 FT DOMAIN 668  
 FT DOMAIN 669  
 FT DOMAIN 670  
 FT DOMAIN 671  
 FT DOMAIN 672  
 FT DOMAIN 673  
 FT DOMAIN 674  
 FT DOMAIN 675  
 FT DOMAIN 676  
 FT DOMAIN 677  
 FT DOMAIN 678  
 FT DOMAIN 679  
 FT DOMAIN 680  
 FT DOMAIN 681  
 FT DOMAIN 682  
 FT DOMAIN 683  
 FT DOMAIN 684  
 FT DOMAIN 685  
 FT DOMAIN 686  
 FT DOMAIN 687  
 FT DOMAIN 688  
 FT DOMAIN 689  
 FT DOMAIN 690  
 FT DOMAIN 691  
 FT DOMAIN 692  
 FT DOMAIN 693  
 FT DOMAIN 694  
 FT DOMAIN 695  
 FT DOMAIN 696  
 FT DOMAIN 697  
 FT DOMAIN 698  
 FT DOMAIN 699  
 FT DOMAIN 700  
 FT DOMAIN 701  
 FT DOMAIN 702  
 FT DOMAIN 703  
 FT DOMAIN 704  
 FT DOMAIN 705  
 FT DOMAIN 706  
 FT DOMAIN 707  
 FT DOMAIN 708  
 FT DOMAIN 709  
 FT DOMAIN 710  
 FT DOMAIN 711  
 FT DOMAIN 712  
 FT DOMAIN 713  
 FT DOMAIN 714  
 FT DOMAIN 715  
 FT DOMAIN 716  
 FT DOMAIN 717  
 FT DOMAIN 718  
 FT DOMAIN 719  
 FT DOMAIN 720  
 FT DOMAIN 721  
 FT DOMAIN 722  
 FT DOMAIN 723  
 FT DOMAIN 724  
 FT DOMAIN 725  
 FT DOMAIN 726  
 FT DOMAIN 727  
 FT DOMAIN 728  
 FT DOMAIN 729  
 FT DOMAIN 730  
 FT DOMAIN 731  
 FT DOMAIN 732  
 FT DOMAIN 733  
 FT DOMAIN 734  
 FT DOMAIN 735  
 FT DOMAIN 736  
 FT DOMAIN 737  
 FT DOMAIN 738  
 FT DOMAIN 739  
 FT DOMAIN 740  
 FT DOMAIN 741  
 FT DOMAIN 742  
 FT DOMAIN 743  
 FT DOMAIN 744  
 FT DOMAIN 745  
 FT DOMAIN 746  
 FT DOMAIN 747  
 FT DOMAIN 748  
 FT DOMAIN 749  
 FT DOMAIN 750  
 FT DOMAIN 751  
 FT DOMAIN 752  
 FT DOMAIN 753  
 FT DOMAIN 754  
 FT DOMAIN 755  
 FT DOMAIN 756  
 FT DOMAIN 757  
 FT DOMAIN 758  
 FT DOMAIN 759  
 FT DOMAIN 760  
 FT DOMAIN 761  
 FT DOMAIN 762  
 FT DOMAIN 763  
 FT DOMAIN 764  
 FT DOMAIN 765  
 FT DOMAIN 766  
 FT DOMAIN 767  
 FT DOMAIN 768  
 FT DOMAIN 769  
 FT DOMAIN 770  
 FT DOMAIN 771  
 FT DOMAIN 772  
 FT DOMAIN 773  
 FT DOMAIN 774  
 FT DOMAIN 775  
 FT DOMAIN 776  
 FT DOMAIN 777  
 FT DOMAIN 778  
 FT DOMAIN 779  
 FT DOMAIN 780  
 FT DOMAIN 781  
 FT DOMAIN 782  
 FT DOMAIN 783  
 FT DOMAIN 784  
 FT DOMAIN 785  
 FT DOMAIN 786  
 FT DOMAIN 787  
 FT DOMAIN 788  
 FT DOMAIN 789  
 FT DOMAIN 790  
 FT DOMAIN 791  
 FT DOMAIN 792  
 FT DOMAIN 793  
 FT DOMAIN 794  
 FT DOMAIN 795  
 FT DOMAIN 796  
 FT DOMAIN 797  
 FT DOMAIN 798  
 FT DOMAIN 799  
 FT DOMAIN 800  
 FT DOMAIN 801  
 FT DOMAIN 802  
 FT DOMAIN 803  
 FT DOMAIN 804  
 FT DOMAIN 805  
 FT DOMAIN 806  
 FT DOMAIN 807  
 FT DOMAIN 808  
 FT DOMAIN 809  
 FT DOMAIN 810  
 FT DOMAIN 811  
 FT DOMAIN 812  
 FT DOMAIN 813  
 FT DOMAIN 814  
 FT DOMAIN 815  
 FT DOMAIN 816  
 FT DOMAIN 817  
 FT DOMAIN 818  
 FT DOMAIN 819  
 FT DOMAIN 820  
 FT DOMAIN 821  
 FT DOMAIN 822  
 FT DOMAIN 823  
 FT DOMAIN 824  
 FT DOMAIN 825  
 FT DOMAIN 826  
 FT DOMAIN 827  
 FT DOMAIN 828  
 FT DOMAIN 829  
 FT DOMAIN 830  
 FT DOMAIN 831  
 FT DOMAIN 832  
 FT DOMAIN 833  
 FT DOMAIN 834  
 FT DOMAIN 835  
 FT DOMAIN 836  
 FT DOMAIN 837  
 FT DOMAIN 838  
 FT DOMAIN 839  
 FT DOMAIN 840  
 FT DOMAIN 841  
 FT DOMAIN 842  
 FT DOMAIN 843  
 FT DOMAIN 844  
 FT DOMAIN 845  
 FT DOMAIN 846  
 FT DOMAIN 847  
 FT DOMAIN 848  
 FT DOMAIN 849  
 FT DOMAIN 850  
 FT DOMAIN 851  
 FT DOMAIN 852  
 FT DOMAIN 853  
 FT DOMAIN 854  
 FT DOMAIN 855  
 FT DOMAIN 856  
 FT DOMAIN 857  
 FT DOMAIN 858  
 FT DOMAIN 859  
 FT DOMAIN 860  
 FT DOMAIN 861  
 FT DOMAIN 862  
 FT DOMAIN 863  
 FT DOMAIN 864  
 FT DOMAIN 865  
 FT DOMAIN 866  
 FT DOMAIN 867  
 FT DOMAIN 868  
 FT DOMAIN 869  
 FT DOMAIN 870  
 FT DOMAIN 871  
 FT DOMAIN 872  
 FT DOMAIN 873  
 FT DOMAIN 874  
 FT DOMAIN 875  
 FT DOMAIN 876  
 FT DOMAIN 877  
 FT DOMAIN 878  
 FT DOMAIN 879  
 FT DOMAIN 880  
 FT DOMAIN 881  
 FT DOMAIN 882  
 FT DOMAIN 883  
 FT DOMAIN 884  
 FT DOMAIN 885  
 FT DOMAIN 886  
 FT DOMAIN 887  
 FT DOMAIN 888  
 FT DOMAIN 889  
 FT DOMAIN 890  
 FT DOMAIN 891  
 FT DOMAIN 892  
 FT DOMAIN 893  
 FT DOMAIN 894  
 FT DOMAIN 895  
 FT DOMAIN 896  
 FT DOMAIN 897  
 FT DOMAIN 898  
 FT DOMAIN 899  
 FT DOMAIN 900  
 FT DOMAIN 901  
 FT DOMAIN 902  
 FT DOMAIN 903  
 FT DOMAIN 904  
 FT DOMAIN 905  
 FT DOMAIN 906  
 FT DOMAIN 907  
 FT DOMAIN 908  
 FT DOMAIN 909  
 FT DOMAIN 910  
 FT DOMAIN 911  
 FT DOMAIN 912  
 FT DOMAIN 913  
 FT DOMAIN 914  
 FT DOMAIN 915  
 FT DOMAIN 916  
 FT DOMAIN 917  
 FT DOMAIN 918  
 FT DOMAIN 919  
 FT DOMAIN 920  
 FT DOMAIN 921  
 FT DOMAIN 922  
 FT DOMAIN 923  
 FT DOMAIN 924  
 FT DOMAIN 925  
 FT DOMAIN 926  
 FT DOMAIN 927  
 FT DOMAIN 928  
 FT DOMAIN 929  
 FT DOMAIN 930  
 FT DOMAIN 931  
 FT DOMAIN 932  
 FT DOMAIN 933  
 FT DOMAIN 934  
 FT DOMAIN 935  
 FT DOMAIN 936  
 FT DOMAIN 937  
 FT DOMAIN 938  
 FT DOMAIN 939  
 FT DOMAIN 940  
 FT DOMAIN 941  
 FT DOMAIN 942  
 FT DOMAIN 943  
 FT DOMAIN 944  
 FT DOMAIN 945  
 FT DOMAIN 946  
 FT DOMAIN 947  
 FT DOMAIN 948  
 FT DOMAIN 949  
 FT DOMAIN 950  
 FT DOMAIN 951  
 FT DOMAIN 952  
 FT DOMAIN 953  
 FT DOMAIN 954  
 FT DOMAIN 955  
 FT DOMAIN 956  
 FT DOMAIN 957  
 FT DOMAIN 958  
 FT DOMAIN 959  
 FT DOMAIN 960  
 FT DOMAIN 961  
 FT DOMAIN 962  
 FT DOMAIN 963  
 FT DOMAIN 964  
 FT DOMAIN 965  
 FT DOMAIN 966  
 FT DOMAIN 967  
 FT DOMAIN 968  
 FT DOMAIN 969  
 FT

	Query Match Similarity	2.6%	Score 101;	DB 1;	Length 3519;	
	Best Local Similarity	19.3%	Pred. No. 12e+02;			
	Matches 122;	Conservative	63;	Mismatches 234;	IndeIs 214;	Gaps 26

  

OY	4	VAVAAVGVLLLAGAGGAAGDEAREAAARVALVARLGGPAPADSVSVERALAKPGID	63
Dd	664	IAAAACVAGGLSLIA-----DGARVVULRSRAIARIAGGCVSVSLPAGRFRRTLEEFD	716
OY	64	TYSLGCGGAAARVRVGSTGVAAAGLHRLRIFCCGH-----VAMSGSQLRLRP	113
Dd	717	-----GRLSVAAVGPSSTVISGVQALDELIIAGCEBEGVARARVPDYASHASQMQL	770
OY	114	LPAVBGEITTEATPNRYRYQNVCTOSYSFVMWDW-----ARNEREID	155
Dd	771	RDEILEALADIPT-----QDSVSFFGSFTVDMLGTITALGAGWFTNLRETRFOEAVE	824
OY	156	WMALNGINILALMSQG-----EAIMQR--VIALGIGTQMEINEFFTGFPAFLMG	202
Dd	825	GIYAQGMGAFWECSPHPVLVPIGETLDLDDNAAFVGLRRBDGDLRFILTS-----	877
OY	203	RMGNLHTWDGPLPSMHLKOLYLCHRVLDOXRSFGMTPLPAFAGHVBEAVTRPEOVNV	262
Dd	878	-----LAFAVGPVPVDMXR-----AFEGVTPTRYVD--LPITYPF	909
OY	263	TKMSGWGHPNCSYSCSFLLARPDPFPILIGSLFLRELKEGTDTHTYGADTTENMQPPSS	322
Dd	910	QROHYW-----LMAEEAPV-----SQPHHS	929
OY	323	EPSYLAATAATTAYEA--MTAVDTAV-----WLQGLWFQHCPQF--WGPAQIRAV	369
Dd	930	ENSFISVAADADAEBAAELGVDVAVEAMPALSSWRQSGLBAEVNQMRDYAMKRILT	989
OY	370	LGAVRP--GRLLYL-----DLFASQPYRTITASFC-QGPIFNCMLENFGNHGLFG-A	419
Dd	990	TGAIPBEKGNNLVLPACTDTTFABS-----LAFTAAAEIGVSYSFAQVDTJAHPRDSQYAA	1046
OY	420	LEAANGGPE-----AARLFENSTWGTGMAPEG-----IS	449
Dd	1047	LROLTGRENVDHLVSLALDQATDLDLAAPBSCLASLYLVAQVLDLRVGBGRILMLVT	1100
OY	450	ONEVVYSLMAELGNRKDPYPDLAAWYTSFPAARVYSHPDAGAANRLILRSYTNCSGEAC	509
Dd	1107	RGAAVAG-PSDAGAVIDPV---QAQVWGCF-GRVGLGLEHEP--IMGGLIDLFPVGDEBYVC	1155
OY	510	R-----GHNRSPLVRPSLQNTNSTWYWR	533
Dd	1159	RRFVGVVASAGEEDQAVR-----GSGGVWRK	1185

  

RESULT 13

ID	CATA_SALTI	STANDARD;	PRT;	726 AA.
AC	082303;			
DT	15-SEP-2003 (Rel. 42; Last Created)			
DT	15-SEP-2003 (Rel. 42; Last Sequence Update)			
DT	15-SEP-2003 (Rel. 42; Last Annotation Update)			
DE	Peroxiase/catalase HPI (EC 1.11.1.6) (Catalase-peroxiase)			
DE	(Hydroperoxiase I).			
GN	KATG OR STY3760 OR T3510.			
OS	Salmonella typhi.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OK	NCBI_Taxid=601;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mouall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Basham D., Brooks K., Chillingworth T., Compton P., Croxin A., Davis P., Davies R.M., Dow L.T., White N., Farrar J., Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,			

R	A	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
R	A	Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
R	A	Whitehead S., Barrall B.G.,
RT	"Complete genome sequence of a multiple drug resistant Salmonella	
RT	enterica serovar Typhimurium CT18."	
RL	Nature 413:848-852(2001).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Ty2 / ATCC 700931;	
RX	MEDLINE=22531367; PubMed=12644504;	
RA	Deng W., Liu S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,	
RA	Burkhard V., Kodoyianni V., Schwartz D.C., Blattner F.R.;	
RT	"Comparative genomics of Salmonella enterica serovar Typhimurium strains Ty2	
RT	and CT18";	
RL	J. Bacteriol. 185:2330-2337(2003).	
CC	- FUNCTION: Bifunctional, exhibiting both a catalase and	
CC	broad-spectrum peroxidase activities (By similarity).	
CC	- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.	
CC	- COFACTOR: Binds 2 prothemin IX and 2 iron ions per tetramer (By	
CC	similarity).	
CC	- SUBUNIT: Homotetramer (By similarity).	
CC	- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL	
CC	PEROXIDASE/CATALASE SUBFAMILY.	
CC	- SIMILARITY: Contains 1 heme regulatory motif (HRM) repeat.	
CC	-----	
CC	This SWISS-PROT entry is copyrighted. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its use by	
CC	non-profit institutions as long as its content is in no way used for commercial	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AL627279; CAD09515.1; -	
DR	EMBL; AE016846; AAC71018.1; -	
DR	InterPro; IPR000763; Bac_citase/pyrase.	
DR	InterPro; IPR002016; Peroxidase.	
DR	Pfam; PF00141; Peroxidase; 1.	
DR	Pfam; PR00458; PEROXIDASE.	
DR	TIGRFAMs; TIGR00198; cat_per_HPI; 1.	
DR	PROSITE; PS00435; PEROXIDASE_1; 1.	
DR	PROSITE; PS00436; PEROXIDASE_2; 1.	
KW	Oxidoreductases; Peroxidase; Iron; Heme; Hydrogen peroxide;	
KW	Complete proteome.	
FT	ACT SITE 102 BY SIMILARITY.	
FT	ACT SITE 106 BY SIMILARITY.	
FT	ACT SITE 106 METAL IRON (HEME AXIAL LIGAND).	
FT	REPEAT 15 20 HRM.	
SO	SEQUENCE 726 AA; 79597 MW; B701C95343EEBB38 CRC64;	
Query Match	2.5%; Score 100; DB 1; Length 726;	
Best Local Similarity	17.6%; Pred. No. 17;	
Matches 133; Conservative	73; Mismatches 197; Indels 352; Gaps 33;	
QY	86 AAGLRHYLRDPCGCHVAVMSGSLRPPLPAVPPELTREATENR-----RRYYQNVTQT 138	
DB	26 SAGAGTASRD-----WRPQLV-----DLNQHNRNSNPGLGEDDYDKKEPSKL 69	
QY	139 SYSFV-----WMDAWEERIRDMALNGINLAALS----- 169	
DB	70 DYSAIKGDALKALTDSOPWW-----PADWGSYVGFLFRAMWAGATYRSIDRGSGAGR 122	
QY	170 -----GQAIMGRVYLALGLTCAEINEFPGTAF 198	
DB	123 GQGTFAPINSPPDVNSLDKARLLMPFKOKTKISWADFLTAGNAVALENSGRT---- 178	
QY	199 LAWMGMGLATWDGPLRPSWHIKOLYLQHRVLDMQMSFGMTPVLPAAQHAVPEAVTRYEP 258	
DB	179 FGFG-AGEDVEDWBDLDVNWDEDAKWLTHR-----H-PEALAKA-- 215	
QY	259 QVNTKXGSMWFHFPGSCSYCSFLIAPDEPIPIIIISLFRELKEFGTGTHIVGADPTENMQ 318	
DB	216 PLGATEKG-----LIYVPEEG-----DH----- 234	



QY 319 PPSSEPSYLAATTAAYEANTAVDTAVALLOGMLFQHPQFQWGPQAQIATVIGAVPRGL 378  
 DB 235 --SGEPLSAAATATATGNNMGNDEEVALIAG--GHTLGKTHGAAASHVQADP-- 285  
 QY 379 LVLDLFAESQPVYTRTSFGQGFPIWOMHNFQGNHGLGALBAVNGGEAA----- 430  
 DB 286 -----EAAPI-----EAGLGMASSYSG-----VGA-DATISGLVAVWTQTPTOM 325  
 QY 431 -----RLPNSMTVGTGAPEGISQNEVVS-----LMAELGWRKD 466  
 DB 326 SNVFEFLFKYEWV--QTRSPAGALQPEAVDADIIIPDPDPKRRKPTMLVTDLFLRD 383  
 QY 467 P-----VPLLAAMVTSFAARRYGVSHPDGAARLLRLRYVNCSEAGCRGHRSL 517  
 DB 384 PEFKISRRLNPQANEAFAFAWFLTHRDMPARFY----- 423  
 QY 518 VRPSLOMNTSIWNRSQVDEAWRLILTSAPSLATSPARVYDLDD-----TROAVDEL 571  
 DB 424 --GPEVKEKDLIWD-----PLPQPLVQPT-QEDILNKAAIASGLSISM 467  
 QY 572 VSLYEEFARSAYLSKELASLLRAG-----GVLAYELLPALDEVL 610  
 DB 468 VSV-----AMAS--ASTFRGDKRGANGARLALAPQRDWEVNAVARVLPVLEALQ 517  
 QY 611 ASDSRFLGSS-----WLEQARAAA-----VSEAEAD 636  
 DB 518 KITNKASLADIYLAGVVGIEQAAAAGASISVFPAPGRVDARQDQDLEMSLEPIAD 577  
 QY 637 FYEQ-----NSRYQLTLMGPEGNII 656  
 DB 578 GFRNVARLDVSTESILIDKAOQLTLTAPEMTVL 612  
 RESULT 14  
 ID ENV\_HV2D1 STANDARD; PRT; 851 AA.  
 AC P17755;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Envelope polypeptide GP160 precursor [contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].  
 GN ENV.  
 OS Human immunodeficiency virus type 2 (Isolate D194) (HIV-2).  
 OC Viruses; Retroviruses; Retroviridae; Lentiviruses.  
 OX NCBI\_TaxID=11713;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91045094; PubMed=2235509;  
 RA Kuehnel H., Kreutz R., Ruebsamen-Waigmann H.;  
 RT "Nucleotide sequence of HIV-2D194, an isolate from a Gambian case of 'neuro-AIDS', which showed excellent growth in macrophages."; Nucleic Acids Res. 18:6142-6142(1990).  
 RL [2]  
 RP SEQUENCE OF 1-266 FROM N.A.  
 RX MEDLINE=9104631; PubMed=2467304;  
 RA Kuehnel H., von Briesen H., Dietrich U., Adamski M., Mix D., Blesert L., Kreutz R., Immanuel A., Henco K., Weichner C., Rasmussen R., Gelderblom H., Ruebsamen-Waigmann H.;  
 RT "Molecular cloning of two west African human immunodeficiency virus type 2 isolates that replicate well in macrophages: a Gambian isolate, from a patient with neurologic acquired immunodeficiency syndrome, and a highly divergent Ghanaian isolate."; Proc. Natl. Acad. Sci. U.S.A. 86:2383-2387(1989).  
 RL [1]  
 CC MISCELLANEOUS: THIS ISOLATE IS FROM A GAMBIAN CASE OF 'NEURO-AIDS'.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: J04542; AA076847.1; -;  
 DR EMBL: X52223; CA36471.1; -;  
 DR PIR: S12159; S12159.  
 DR HIV: J04542; Env: GP41.  
 DR InterPro: IPR000328; Env: GP41.  
 DR InterPro: IPR000777; GP120.  
 DR Pfam: PF00517; GP41; 1.  
 DR AIDS: Coat protein; Polypeptide; Glycoprotein; Transmembrane;  
 KW Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 501  
 FT CHAIN 502 851  
 FT CARBOHYD 37 37  
 FT CARBOHYD 70 70  
 FT CARBOHYD 114 114  
 FT CARBOHYD 127 127  
 FT CARBOHYD 134 134  
 FT CARBOHYD 142 142  
 FT CARBOHYD 157 157  
 FT CARBOHYD 184 184  
 FT CARBOHYD 195 195  
 FT CARBOHYD 227 227  
 FT CARBOHYD 230 230  
 FT CARBOHYD 261 261  
 FT CARBOHYD 267 267  
 FT CARBOHYD 278 278  
 FT CARBOHYD 289 289  
 FT CARBOHYD 299 299  
 FT CARBOHYD 355 355  
 FT CARBOHYD 361 361  
 FT CARBOHYD 388 388  
 FT CARBOHYD 398 398  
 FT CARBOHYD 401 401  
 FT CARBOHYD 438 438  
 FT CARBOHYD 453 453  
 FT CARBOHYD 456 456  
 FT CARBOHYD 601 601  
 FT CARBOHYD 610 610  
 FT CARBOHYD 626 626  
 SQ SEQUENCE 851 AA; 97178 MW; 3B02FC0B9FF118 CRC64;  
 Query Match 2.5%; Score 100; DB 1; Length 851;  
 Best Local Similarity 20.3%; Pred. No. 21;  
 Matches 105; Conservative 68; Mismatches 185; Indels 160; Gaps 29;  
 QY 292 GSLFLR---ELIPEGTDHYGADTNEWOPES-----SEPSY-----LAAAT 331  
 DB 187 GTCWRHONTSVIVESCDKHWDMMKFRYCAPPGFALLRCNDTVSGFEPKCSKVVAASC 246  
 QY 332 TAVEAMTAVDTAVALVTL-LOGMLFQHPQ-FWGPQAQIATVIG-----AVPRGL 378  
 DB 247 TRMETQT-----STWGFNGTBAENRTIYMHGDKNRITISLNKYNNLTMHGCRPQNK 301  
 QY 379 LV-LDLFA---ESQPVYTRTSFGQGFPIWOMHNFQGNHGLGALBAVNGGEAAELF 433  
 DB 302 VVPIILSGRFRHSPVYNKK--PQAWCW-----FGNW-----LEAREVQOTLAKH 348  
 QY 434 P-----NSTMTGTGAPEGISQNEVVS-----LMAELGWRKDVPVDLAAW 474  
 DB 349 FRYGNDTGTGKINTKFGISGP-----EYTWMTNCRGSEFLYCNMTW-----FLNW 395  
 QY 475 VTSPARRYGVSHD---DAGAAWRLILRSYV--NCSGE-ACRGHNSPLVRESLOMT 527  
 DB 396 VENKTNGHGVACCHIRQIINTWIKVGTNYVLEPRREGELTNSVTISIILNIDSDNOT 455  
 QY 528 SIWNRSVFPAMFALLTSAPSLATSPAFRYDLIDLROAVDELVSUYEEA-----RSA 582  
 DB 456 NITS-ARVAVLYLELGD-----KLEIVTPIPPAPYKERYSSAPVANKGV 503

QY 583 YLSKELASLLRAGVLAPELLALBDEVLASDSRFLGSGWLEQARAAVSEADPYE-QN 641  
 Db 504 FVLGFGFLATAGSANGASL-----TISAQRRTLAGIVQO-----QQCLLDVKKQO 552  
 QY 642 SRVQLTLWGP-----EGNILDYANKOLAGL-----VANYTTPMR--- 676  
 Db 553 EMLRLTWGCTKQLQARVTAIEKYLKDAQLNMGCAFQVCHTTPVWVNDSLTPMNNMT 612  
 QY 677 -----LFLBALDVSVAQGIPIFGQHDPKXVFOLEQ 706  
 Db 613 WOEWKRVHYLEA---NISQSLQAOAQIQEKKMYELQK 647

RESULT 15  
 NME4\_MOUSE STANDARD; PRT: 1323 AA.  
 ID NME4\_MOUSE  
 AC 003391;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glutamate (NMDA) receptor subunit epsilon 4 precursor (N-methyl)  
 GN D-separate receptor subtype 2D (NR2D) (NMDAR2D).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID:10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93050214; PubMed=1385220;  
 RA Ikeda K., Nagaawa M., Mori H., Araki K., Sakimura K., Watanabe M.,  
 RA Inoue Y., Mishina M.;  
 RT "Cloning and expression of the epsilon 4 subunit of the NMDA receptor  
 RT channel.";  
 RL FEBS Lett. 313:34-38(1992).  
 RN [2]  
 RP REVISIONS.  
 RA Ikeda K., Nagaawa M., Mori H., Araki K., Sakimura K., Watanabe M.,  
 RA Inoue Y., Mishina M.;  
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS  
 CC POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT  
 CC SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.  
 CC -1- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D12822; BAA0254.1; -  
 DR EMBL: P19491; IGR2.  
 DR MGD: MGJ:95823; Grind.  
 DR InterPro: IPR001828; ANF\_receptor.  
 DR InterPro: IPR001330; Ion\_glu\_receptor.  
 DR InterPro: IPR001508; NMDA\_receptor.  
 DR InterPro: IPR001311; SBP\_glu\_receptor.  
 DR Pfam: PF01094; ANF\_receptor; 1.  
 DR Pfam: PF00060; 1;g\_chan; 1.  
 DR PRINTS: PR00177; NMDARECEPTOR.  
 DR SMART: SM00079; pape; 1.  
 KM Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;  
 KM Ionic channel; Magnesium.  
 FT SIGNAL 1 27  
 FT CHAIN 28 1323  
 FT DOMAIN 28 580  
 FT EXTRACELLULAR (POTENTIAL).

FT TRANSEM 581 601  
 FT DOMAIN 602 623  
 FT TRANSEM 624 644  
 FT DOMAIN 645 653  
 FT TRANSEM 654 674  
 FT DOMAIN 675 681  
 FT TRANSEM 842 862  
 FT DOMAIN 863 1323  
 FT DOMAIN 278 283  
 FT DOMAIN 905 913  
 FT DOMAIN 1030 1035  
 FT DOMAIN 1197 1201  
 FT SITE 639 639

FT CARBOHYD 89 89  
 FT CARBOHYD 349 349  
 FT CARBOHYD 363 363  
 FT CARBOHYD 464 464  
 FT CARBOHYD 566 566  
 SQ SEQUENCE 1323 AA; 142907 MM; 8AE9878F90DD0921 CRC64;

Query Match 2.58; Score 100; DB 1; Length 1323;  
 Best Local Similarity 22.78; Pred. No. 39;  
 Matches 106; Conservative 46; Mismatches 156; Indels 158; Gaps 28;

QY 162 INITIALMSGQGEATWQRYVYALGLTOAEINEFFTGPAFLAWGKGNLH-----TWQGLP 215  
 Db 48 LNVAVLFSGP-----AYAAEARL-----GPAVAAVRSPLGDIVRVALVINGSDP 93  
 QY 216 PSHIRKQLYIQ-----HRYL--DQMSFGMPVLPAPAGHYPEAVTRVFPQVNT 263  
 Db 94 -----RSLVQLQDLISGLRVHGVFEDDSRAVAVPIIDFLSAQTSLPVAVHG----- 143  
 QY 264 KMGSMGHFNGSYSCSFLLAPEDPIPTIGSLFRELIKERTGHTIGADTFNEMO----- 318  
 Db 144 -----GALVLTPEK-----GSTFLQ-----LGSIEQQLQVFEV 175  
 QY 319 -PSESPSYLAATTA-----VYEMTAVDTEAVMLQGMLEFOHOPF-WGPAQIRAVLGA 372  
 Db 176 LEEYDWTSPVAVTTAPGRHRAFLSYIEVLTDG--SLVGW--EHRGALTLDPGAGAVLGA 231  
 QY 373 VPRG-----RLVLDLFAESQPIY--TRTASFOGQAPPIWGMH--NFGNHGLFG-ALE 421  
 Db 232 QLRVSAQIRLLFC-AEEZAEVFERAAEEAGLTPGVWPMVGPQLAGGGSGVPGEPPL 290  
 QY 422 AVNGGPEAARLFPNSTWGTGMAPEGISQNEVVYSIMAEIGWRKDPVPDIAAVTSPA-- 479  
 Db 291 LPGAPLPAGLF-----AVRSAGWRDDLARRVAAGVAVVARG 327  
 QY 480 ---ARRYGVSHPDAGAMRLLIRSVYNGSGEACRGHNSPLVRRPQLQ--NNTSIWYN 532  
 Db 328 AQAALIRYGFLL-DEL-----GHQCPAQRNRTH--RGESLHRYFPANIT-WDN 368  
 QY 533 RSDVFEAMRLLTSPASLATSPPAFRYDLDTTQAOVELSLTYEE 578  
 Db 369 RQVSEFNEDEFLVN--PSLV-----VISLTDRTWEVVGSMEOQ 404

Search completed: February 13, 2004, 16:15:50  
 Job time : 26 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2004, 15:59:34 ; Search time 61 Seconds

(Without alignments)  
3143.165 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939

Sequence: 1 MEAVVAAGVLLAGAG.....VDLAKKIFLKYPGWAGSW 743

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaea:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3306	83.9	639	4 Q14769	Q14769 Homo sapien
2	3305	83.9	739	11 O54752	O54752 mus musculu
3	3302	83.8	739	11 O88325	O88325 mus musculu
4	2613.5	66.3	753	13 Q90276	Q90276 dromais no
5	2611.5	66.3	753	13 Q90275	Q90275 dromais no
6	1522.5	38.7	806	10 Q9FN43	Q9FN43 arabidopsis
7	1445	36.7	811	10 Q9ZRA5	Q9ZRA5 nicotiana t
8	1433.5	36.4	778	5 Q9VLL5	Q9VLL5 diosiphila
9	1060.5	26.9	770	16 Q9AA06	Q9AA06 caulobacter
10	1027.5	26.1	715	5 Q9NAP6	Q9NAP6 caenorhabdi
11	962	24.4	799	16 Q8P87	Q8P87 xanthomonas
12	854	21.7	2104	16 Q8XM24	Q8XM24 clostridium
13	143.5	3.6	1862	2 Q93TW7	Q93TW7 eligmateila
14	128	3.2	444	3 Q96W93	Q96W93 gibberella
15	124.5	3.2	783	16 Q9FBV4	Q9FBV4 streptomyces
16	122.5	3.1	3247	12 Q65553	Q65553 bovine hept

17	121	3.1	2116	2 Q93N89	Q93N89 streptomyces
18	121	3.1	4342	16 Q91157	Q91157 pseudomonas
19	121	3.1	4472	2 O33954	O33954 streptomyces
20	120	3.0	3729	2 O33956	O33956 streptomyces
21	118	3.0	2785	17 Q8ZYB9	Q8ZYB9 pyrobaculum
22	117.5	3.0	1490	17 Q8PW14	Q8PW14 methanosarc
23	117	3.0	1374	17 Q9YCP5	Q9YCP5 aeropyrum p
24	116	2.9	1795	2 Q8KPM5	Q8KPM5 streptomyces
25	115.5	2.9	973	16 Q86728	Q86728 streptomyces
26	115.5	2.9	1319	4 Q9HCD3	Q9HCD3 homo sapien
27	115	2.9	725	16 Q9A722	Q9A722 caulobacter
28	115	2.9	7257	2 Q9K127	Q9K127 polyangium
29	114.5	2.9	2338	16 Q8X119	Q8X119 ralsionia s
30	114	2.9	479	2 Q86001	Q86001 sphingomonas
31	113.5	2.9	1030	16 Q8P655	Q8P655 xanthomonas
32	113.5	2.9	2164	2 Q93N86	Q93N86 streptomyces
33	113	2.9	1281	16 Q91211	Q91211 pseudomonas
34	113	2.9	7257	2 Q916C7	Q916C7 polyangium
35	112.5	2.9	380	16 Q8P1D5	Q8P1D5 xanthomonas
36	112.5	2.9	839	16 Q8P3W4	Q8P3W4 xanthomonas
37	112	2.8	596	2 Q9R8J7	Q9R8J7 xanthomonas
38	112	2.8	768	3 Q8J274	Q8J274 grifolia fro
39	111.5	2.8	484	10 Q9FFZ1	Q9FFZ1 arabidopsis
40	111.5	2.8	689	10 Q9AMX9	Q9AMX9 oryza sativ
41	111.5	2.8	2205	12 Q99FK2	Q99FK2 porcine tes
42	111	2.8	1027	16 Q9AA01	Q9AA01 caulobacter
43	111	2.8	10917	2 Q93NM6	Q93NM6 streptomyces
44	110.5	2.8	582	2 Q917A1	Q917A1 streptomyces
45	110.5	2.8	614	2 Q916K3	Q916K3 mycobacteri

## ALIGNMENTS

RESULT 1  
ID Q14769 PRELIMINARY; PRT; 639 AA.  
AC Q14769;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE Alpha-N-acetylglucosaminidase.  
GN UPHSD2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96359157; PubMed=8703123;  
RA Zhao Z., Yazdani A., Shen Y., Sun Z., Bailey J., Caskey C.T.,  
Lee C.C.;  
RT "Molecular dissection of a cosmid from a gene-rich region in 17q21 and  
RT characterization of a candidate gene for alpha-N-acetylglucosaminidase  
RT with two cDNA isoforms".  
RL Mamm. Genome 7:686-690 (1996).  
DR EMBL; U78455; AAB36505.1; -  
SQ SEQUENCE 639 AA; 72027 MW; 21847A1A2E9786 CRC64;

Query Match 83.9%; Score 3306; DB 4; Length 639;  
Best Local Similarity 99.7%; Pred. No. 3.5e-233;  
Matches 617; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	125	TPNRYRYQNTQSYSPVWMDMARERIDMVALNGINLALAMSGQEAIMQRYVIAIGL	184
DB	21	TARRRYQNTQSYSPVWMDMARERIDMVALNGINLALAMSGQEAIMQRYVIAIGL	80
QY	185	TAQINEFTGPAFLAMGNGNLHTMDGFLPSWIKIOLYQHRVLDQMSGMPVPA	244
DB	81	TAQINEFTGPAFLAMGNGNLHTMDGFLPSWIKIOLYQHRVLDQMSGMPVPA	140
QY	245	PAGHYEAVTRVFPQVNTYKMGWGFNCYSCEFLAPEDDIPFIIGSLFLRELKBERG	304

```

Db 141 FASHVPEAVTRVPEQVNVTKGSGWGHFNCSYCSFLAPEDPIFPIIGSLFIRELIXEKG 200
Qy 305 TCHYAGDNTENEMQPPSSPSYLAATAATYAEAMTAVTEAVVLLQGMFLFOHOPQWGA 364
Db 201 TDHLYGADTNEHQPPSSPSYLAATAATYAEAMTAVTEAVVLLQGMFLFOHOPQWGA 260
Qy 365 QIRAVLCAVPRGRLLVLDLPAESQPVYTRTASFQGGPFTWCMHNFQGNHGLFGALEAVN 424
Db 261 QIRAVLCAVPRGRLLVLDLPAESQPVYTRTASFQGGPFTWCMHNFQGNHGLFGALEAVN 320
Qy 425 GGPEAAFLFNSNTWVGMAPEGISONEVYYSIMAEIGMRKDPVPLAAVWTSFAARVY 484
Db 321 GGPEAAFLFNSNTWVGMAPEGISONEVYYSIMAEIGMRKDPVPLAAVWTSFAARVY 380
Qy 485 VSHPDACAARLLLSRYNCSGEACGHNRSPLVRPSLQMTSITWYNSDYVEAVRLL 544
Db 381 VSHPDACAARLLLSRYNCSGEACGHNRSPLVRPSLQMTSITWYNSDYVEAVRLL 440
Qy 545 TSAPSLATSPAFRDLIDLTRQAVQELVSLYEBAASAVLSKEIASLRAAGVLAELLP 604
Db 441 TSAPSLATSPAFRDLIDLTRQAVQELVSLYEBAASAVLSKEIASLRAAGVLAELLP 500
Qy 605 ALDEVILASDSRFLIGSWLEQARAAYSEAEADFEQNSRYQLTLWGPEGNIIDYANKOLA 664
Db 501 ALDEVILASDSRFLIGSWLEQARAAYSEAEADFEQNSRYQLTLWGPEGNIIDYANKOLA 560
Qy 665 GLVANYTTPWRMLFLELVVSVAGQIFPQOHQFDKXVFQLEQAFVLSKQRYPSQPRGDTV 724
Db 561 GLVANYTTPWRMLFLELVVSVAGQIFPQOHQFDKXVFQLEQAFVLSKQRYPSQPRGDTV 620
Qy 725 DLAKKIFLKTYPGWVAGSW 743
Db 621 DLAKKIFLKTYPGWVAGSW 639

RESULT 2
054752 PRELIMINARY; PRT; 739 AA.
ID 054752;
AC 054752;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Naglu (EC 3.2.1.50).
GN NAGLU.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAF1J; TISSUE=Liver;
RA Zhao K.M., Li H.H., Neufeld E.F.;
RT "Cloning and expression of Mouse gene encoding the lysosomal alpha-N-
RT acetylglucosaminidase.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U85247; AAB88084.1; -.
DR MGD; MGI:135164; Naglu.
DR InterPro; IPR002086; Aldenhyde dehydr.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
KM Glycosidase; Hydrolase.
SQ SEQUENCE 739 AA; 82610 MM; B414C336A04EFOEF CRC64;

Query Match 83.9%; Score 3305; DB 11; Length 739;
Best Local Similarity 83.0%; Pred. No. 5,2e-233;
Matches 611; Conservative 53; Mismatches 70; Indels 2; Gaps 1;

```

```

Qy 121 LTEATPNKRYRYQNVQTSYSFVWMDMAREREIDMNLNGINLALMSGQEAIMQRYVL 180
Db 119 LRETPNKRYYQNVQTSYSFVWMDMAREREIDMNLNGINLALMSGQEAIMQRYVL 178
Qy 181 ALGTLQAEINFEFFGPAFLANGWGNHTWDGFLPESNHLKQLYLQHRVLDQNSFGMT 240
Db 179 ALGTLQAEINFEFFGPAFLANGWGNHTWDGFLPESNHLKQLYLQHRVLDQNSFGMT 238
Qy 241 VLPAPAGVPEAVTRVPEQVNVTKGSGWGHFNCSYCSFLAPEDPIFPIIGSLFIRELI 300
Db 239 VLPAPAGVPEAVTRVPEQVNVTKGSGWGHFNCSYCSFLAPEDPIFPIIGSLFIRELI 298
Qy 301 KEFGTDHYGADTNEHQPPSSPSYLAATAATYAEAMTAVTEAVVLLQGMFLFOHOPQ 360
Db 299 KEFGTDHYGADTNEHQPPSSPSYLAATAATYAEAMTAVTEAVVLLQGMFLFOHOPQ 358
Qy 361 MGPAQIRAVLCAVPRGRLLVLDLPAESQPVYTRTASFQGGPFTWCMHNFQGNHGLFGAL 420
Db 359 MGPAQIRAVLCAVPRGRLLVLDLPAESQPVYTRTASFQGGPFTWCMHNFQGNHGLFGAL 418
Qy 421 EAVNGPEAAFLFNSNTWVGMAPEGISONEVYYSIMAEIGMRKDPVPLAAVWTSFAA 480
Db 419 EDVNRGPOAAFLFNSNTWVGMAPEGISONEVYYSIMAEIGMRKDPVPLAAVWTSFAA 478
Qy 481 RRYGVSHPDACAARLLLSRYNCSGEACGHNRSPLVRPSLQMTSITWYNSDYVEAVR 540
Db 479 RRYGVSHPDACAARLLLSRYNCSGEACGHNRSPLVRPSLQMTSITWYNSDYVEAVR 538
Qy 541 RLTLTAPSLATSPAFRDLIDLTRQAVQELVSLYEBAASAVLSKEIASLRAAGVLAEL 600
Db 539 RLTLTAPSLATSPAFRDLIDLTRQAVQELVSLYEBAASAVLSKEIASLRAAGVLAEL 598
Qy 601 ELPLALDEVILASDSRFLIGSWLEQARAAYSEAEADFEQNSRYQLTLWGPEGNIIDYAN 660
Db 599 ELPLALDEVILASDSRFLIGSWLEQARAAYSEAEADFEQNSRYQLTLWGPEGNIIDYAN 658
Qy 661 KOLAGLVANYTTPWRMLFLELVVSVAGQIFPQOHQFDKXVFQLEQAFVLSKQRYPSQPR 720
Db 659 KOLAGLVANYTTPWRMLFLELVVSVAGQIFPQOHQFDKXVFQLEQAFVLSKQRYPSQPR 718
Qy 721 GDTVDLAKKIFLKTYPG 736
Db 719 GDTVDLAKKIFLKTYPG 734

RESULT 3
088325 PRELIMINARY; PRT; 739 AA.
ID 088325;
AC 088325;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Alpha-N-acetylglucosaminidase.
GN NAGLU.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RC MEDLINE=20056274; PubMed=10588735;
RA Li H.H., Yu W.H., Rozengurt N., Zhao H.Z., Lyons K.M.,
RA Anagnostaras S., Fausel M.S., Suzuki K., Vanter M.T., Neufeld E.F.;
RT "Mouse model of Sanfilippo syndrome type B produced by targeted
RT disruption of the gene encoding alpha-N-acetylglucosaminidase.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14505-14510(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvEvTacB; TISSUE=Spleen;
RA Glockner C.J., Brechtling R., Moeller G., Adamski J.;
RT "Characterization of HSD17B1 gene in mice.";

```

Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF003255; AAC26842.1; -  
 DR EMBL: AF363242; AAM2194.1; -  
 DR MGI: 1351641; Naglu.  
 DR InterPro: IPR002086; Aldenhyde dehydr.  
 DR PROSITE: PS00070; ALDENHYDE DEHYDR. CY5; 1.  
 DR SEQUENCE 739 AA; 82596 MW; 01C71966557CDE27 CRC64;

Query Match 83.8%; Score 3302; DB 11; Length 739;  
 Best Local Similarity 82.9%; Pred. No. 8.6e-233;  
 Matches 610; Conservative 54; Mismatches 70; Indels 2; Gaps 1;

QY 1 MEA VAAAVGVLLAGAGGAAGDEARAAVRAVLAALGPPAADPSVERALAKP 60  
 DB 1 MEAAGLAVILGFL--AGGSVGDARAKAKAVRELVRLLGPPRAAFVSVERALDES 58  
 QY 61 GLDTYSLGGGGAARVRGSGTGAAGLHRYLRDFCGCHVANGSGQLRPRPLAVGE 120  
 DB 59 GLDTYSLGGGGAARVRGSGTGAAGLHRYLRDFCGCHVANGSGQLRPRPLAVPDG 118  
 QY 121 LTEATENRYRYONCTOSYSPFWMMDMAREREIDMANGINIALMSGGEALWQRYVL 180  
 DB 119 LLETENRYRYONCTOSYSPFWMMDMAREREIDMANGINIALMSGGEALWQRYVL 178  
 QY 181 ALGSLTQAEINEFPGPAFLANGMGNLHTWDGFLPPSWHLKQLYLQHRVLDQMSFGMT 240  
 DB 179 ALGSLTQAEINEFPGPAFLANGMGNLHTWDGFLPPSWHLKQLYLQHRVLDQMSFGMT 238  
 QY 241 VLPAAGVPAARVRVFPQVNVTKMSKGHNCSYSCSFLAPEDLPFLIGSLFLBELT 300  
 DB 239 VLPAAGVPAARVRVFPQVNVTKMSKGHNCSYSCSFLAPEDLPFLIGSLFLBELT 298  
 QY 301 KEFGDTHYAGADTENEMOPSSSEPSYLAATTAAYEAMTAVTEAWLLQGMFLFOHOP 360  
 DB 299 KEFGDTHYAGADTENEMOPSSSEPSYLAATTAAYEAMTAVTEAWLLQGMFLFOHOP 358  
 QY 361 WGPQIRAVLCAVPRGRLLVLDLFAESQPVYTRTASFOGQPFICMLHNFQNGHFLGAL 420  
 DB 359 WGPQIRAVLCAVPRGRLLVLDLFAESQPVYTRTASFOGQPFICMLHNFQNGHFLGAL 418  
 QY 421 EAVNGGPAARLPFNSITWGTGMAPEGISONEVYVSLMAELGMRKDPVPLAAMVTSFAA 480  
 DB 419 EDVNRGPAARLPFNSITWGTGMAPEGISONEVYVSLMAELGMRKDPVPLAAMVTSFAA 478  
 QY 481 RRYGVSHDAGAAARLLRSVYNGSGEACRGNHNSPLVREPSLQNTSITWNSDVEAM 540  
 DB 479 RRYGVSHDAGAAARLLRSVYNGSGEACRGNHNSPLVREPSLQNTSITWNSDVEAM 538  
 QY 541 RLILTSASLATSAPFRYDLDLTROAVQELVSLYEERASATYSKELASILRAGVTLAY 600  
 DB 539 RLILTSASLATSAPFRYDLDLTROAVQELVSLYEERASATYSKELASILRAGVTLAY 598  
 QY 601 ELPLALDEVLASDSRFLGSMLEQARAASVSEADAFYEONSRYQLTLMGEGNILLDAN 660  
 DB 599 ELPLALDEVLASDSRFLGSMLEQARAASVSEADAFYEONSRYQLTLMGEGNILLDAN 658  
 QY 661 KQLAGLVANVTTPKRLFLFLEALVDSVAQGIFFQOHQPDKNVFOLEQFVLSKQRYPSQPR 720  
 DB 659 KQLAGLVANVTTPKRLFLFLEALVDSVAQGIFFQOHQPDKNVFOLEQFVLSKQRYPSQPR 718  
 QY 721 GDTVDLAKKIFLKYP 736  
 DB 719 GDTVDLAKKIFLKYP 734

## RESULT 4

Q90276 PRELIMINARY; PRT; 753 AA.  
 ID Q90276  
 AC Q90276  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
 DE Lysosomal alpha-N-acetyl glucosaminidase.

OS Dromaius novaehollandiae (Emu).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Palaeognathae; Casuariiformes; Dromadidae;  
 OC Dromaius.  
 OX NCBI\_TaxID=8790;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21309063; PubMed=11414757;  
 RA Aronovich E.L.; Johnson J.M.; Wang P.; Giger U.; Whitley C.B.;  
 RT "Molecular Basis of Mucopolysaccharidosis Type IIb in Emu (Dromaius  
 novaehollandiae): An Avian Model of Sanfilippo Syndrome Type B.";  
 RL Genomics 74:299-305(2001).  
 DR EMBL: AF331668; AK73654.1; -  
 DR SEQUENCE 753 AA; 84013 MW; C4B5332A6775EC01 CRC64;

Query Match 66.3%; Score 2613.5; DB 13; Length 753;  
 Best Local Similarity 64.7%; Pred. No. 2e-182;  
 Matches 480; Conservative 105; Mismatches 144; Indels 13; Gaps 6;

QY 4 VAAVAAVGVLLAGAGG-----AAGDEARAAVRAVLAALGPPAADPSVERAL 56  
 DB 5 VRLTALAMMAAARASPLPRVACAEARQDEAAVRAVLAALGPPRAA-VALSVDGL 63  
 QY 57 AAKPGDITYSLGG--GGAARVRGSGTGAAGLHRYLRDFCGCHVANGSGQLRPRPL 114  
 DB 64 AAGDITRYVHSPGAAVAAGSSGVAAAAGLHRYLRDFCGCHLSWGHGQLRPLDPL 122  
 QY 115 PAVGELTEATENRYRYONCTOSYSPFWMMDMAREREIDMANGINIALMSGGEAL 174  
 DB 123 PAVGELTEATENRYRYONCTOSYSPFWMMDMAREREIDMANGINIALMSGGEAL 182  
 QY 175 WQRYVLAALGTLQAEINEFPGPAFLANGMGNLHTWDGFLPPSWHLKQLYLQHRVLDQMR 234  
 DB 183 WQRYVLAALGTLQAEINEFPGPAFLANGMGNLHTWDGFLPPSWHLKQLYLQHRVLDQMR 242  
 QY 235 SFGMTVYLPAGVHPEAVTRVFPQVNVTKMSKGHNCSYSCSFLAPEDLPFLIGSL 294  
 DB 243 SFGMTVYLPAGVHPEAVTRVFPQVNVTKMSKGHNCSYSCSFLAPEDLPFLIGSL 302  
 QY 295 FLRELKEFGDTHYAGADTENEMOPSSSEPSYLAATTAAYEAMTAVTEAWLLQGMFL 354  
 DB 303 FLRELKEFGDTHYAGADTENEMOPSSSEPSYLAATTAAYEAMTAVTEAWLLQGMFL 362  
 QY 355 QHQPQFMPAQIRAVLCAVPRGRLLVLDLFAESQPVYTRTASFOGQPFICMLHNFQNGH 414  
 DB 363 QHQPQFMPAQIRAVLCAVPRGRLLVLDLFAESQPVYTRTASFOGQPFICMLHNFQNGH 422  
 QY 415 GLFGALAEVNGGPAARLPFNSITWGTGMAPEGISONEVYVSLMAELGMRKDPVPLAAM 474  
 DB 423 GLFGALAEVNGGPAARLPFNSITWGTGMAPEGISONEVYVSLMAELGMRKDPVPLAAM 481  
 QY 475 VTSFAARVYGVSHDAGAAARLLRSVYNGSGEACRGNHNSPLVREPSLQNTSITWNS 534  
 DB 482 VTSFAARVYGVSHDAGAAARLLRSVYNGSGEACRGNHNSPLVREPSLQNTSITWNS 540  
 QY 535 DVEEAWRLILTSAPFRYDLDLTROAVQELVSLYEERASATYSKELASILRAGVTLAY 594  
 DB 541 DVEEAWRLILTSAPFRYDLDLTROAVQELVSLYEERASATYSKELASILRAGVTLAY 600  
 QY 595 GGVLYAYELLPALDEVLASDSRFLGSMLEQARAASVSEADAFYEONSRYQLTLMGEGN 654  
 DB 601 GGVLYAYELLPALDEVLASDSRFLGSMLEQARAASVSEADAFYEONSRYQLTLMGEGN 660  
 QY 655 ILDYANKQLAGLVANVTTPKRLFLFLEALVDSVAQGIFFQOHQPDKNVFOLEQFVLSKQRY 714  
 DB 661 ILDYANKQLAGLVANVTTPKRLFLFLEALVDSVAQGIFFQOHQPDKNVFOLEQFVLSKQRY 720  
 QY 715 YPSQPRGDTVDLAKKIFLKYP 736  
 DB 721 YPSQPRGDTVDLAKKIFLKYP 742

## RESULT 5

090275 ID Q90275 PRELIMINARY; PRT; 753 AA.  
AC Q90275;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Lysosomal alpha-N-acetyl glucosaminidase.  
OS Drosophila novae-hollandiae (Emu).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Palaeognathae; Casuariiformes; Dromaiidae;  
OC Dromaius.  
OC NCBI\_TaxID=8790;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21309063; PubMed=11414757;  
RA Aronovich E.L., Johnston J.M., Wang P., Giger U., Whitley C.B.;  
RT "Molecular Basis of Mucopolysaccharidosis Type IIIB in Emu (Dromaius  
novae-hollandiae): An Avian Model of Sanfilippo Syndrome Type B.";  
RL Genomics 74:299-305(2001).  
DR EMBL; AF31669; AAK73655.1; -.  
SQ SEQUENCE 753 AA; 84008 MW; 3B0445B30B48A85 CRC64;

Query Match 66.3%; Score 2611.5; DB 13; Length 753;  
Best Local Similarity 64.7%; Pred. No. 2.8e-182;  
Matches 480; Conservative 104; Mismatches 145; Indels 13; Gaps 6;

QY 4 VAVAAVGVLLLAGAGG-----AAGDEAREAAVRAVAVLGGPAADESVVERAL 56  
DB 5 VRLTALAAWAAARASPLPEYVRAAGEDARCEAAVRLALRLGPRAA-VALSVDSGL 63  
QY 57 AAKRGIDTSLGG--GGAARVRVRGSGTGAAGAAAGHRYRDRFCGCVAMSGSGLRPL 114  
DB 64 AA-GGIDTTRVHPPGAVAVAAGSSGYAAAAGHRYRDRFCGCVAMSGSGLRPL 122  
QY 115 PAVPGELEATPRRYRYQVCTQSYSEVWDMAREREIDWALNGINALAMSGQEAR 174  
DB 123 PRVPAIRATAPGRFRYYQVCTQSYSEVWDMAREREIDWALNGINALAMSGQEAR 182  
QY 175 WQRYVATLGLTQAEINEFFTGPAFLMGRMGNIHTMDGRLPSMHIKQYLOHRYLDMR 234  
DB 183 WQRYVATLGLTQAEINEFFTGPAFLMGRMGNIHTMDGRLPSMHIKQYLOHRYLDMR 242  
QY 235 SPQMTPLPAFAGHVEAVTRVPPQVNVITKMGSGHFNCSYSCSPLADEDIFPLISGL 294  
DB 243 SLGMITPLPAFAGHVEAVTRVPPQVNVITKMGSGHFNCSYSCSPLADEDIFPLISGL 302  
QY 295 FLBELIKEGTDTHTYADPTNEMQPPSSPSYLAATTAAYEMKTVDEAVWLLQGMFL 354  
DB 303 FLKELLKEGTDTHTYADPTNEMQPPSSPSYLAATTAAYEMKTVDEAVWLLQGMFL 362  
QY 355 QHQPWFQPAQIRAVLGAVERGRLLVTLFAESQPYTRTASFOGQPFIMCMHNFQGNH 414  
DB 363 QHQPWFQPAQIRAVLGAVERGRLLVTLFAESQPYTRTASFOGQPFIMCMHNFQGNH 422  
QY 415 GLFGALAVANGGPDAARLFPNSTMGTCNAPEGISQNEVVYSLMELGMRKQDVPDLAAM 474  
DB 423 GLFGALAVANGGPDAARLFPNSTMGTCNAPEGISQNEVVYSLMELGMRKQDVPDLAAM 481  
QY 475 VTSFAARVGVSHPDGAARWLLLRSYNCSGEGAGHNRSPVLRPSLOMTSINYNS 534  
DB 482 VARYAERKRYAPNAAASAKXLLRSYNTCTG-VCTNHNRSPLVRPSLRMDTEVYNTKS 540  
QY 535 DVEAWELLTTSAPSLATSPAFRYDLDTLROAVQELVSLYEBASAVLSKELASLRA 594  
DB 541 DVEAWELLTTSAPSLATSPAFRYDLDTLROAVQELVSLYEBASAVLSKELASLRA 600  
QY 595 GGYAVVLELPALEBVLASDSRFLGSLLEQARAAVSEAEADYEQNSRYOLTLWPEGN 654  
DB 601 GGYAVVLELPALEBVLASDSRFLGSLLEQARAAVSEAEADYEQNSRYOLTLWPEGN 660  
QY 655 ILIDYANKQAGLVANNTYTPRMRFLLEALYDVAQGIPIFOGHQPKNVFQLEQAFVLSKOR 714  
DB 661 ILIDYANKQAGLVANNTYTPRMRFLLEALYDVAQGIPIFOGHQPKNVFQLEQAFVLSKOR 720

QY 715 YPSQPGDVTDLAKTIFLKTYP 736  
DB 721 YPTAPVGDTEIFISKITFLKTYP 742

RESULT 6  
ID Q9FNA3 PRELIMINARY; PRT; 806 AA.  
AC Q9FNA3;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Putative alpha-N-acetylglucosaminidase.  
GN AT5G13690.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC Eustroideae; Brassicales; Brassicaceae; Arabidopsids.  
OC NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Columbia;  
RY MEDLINE=98069011; PubMed=9405937;  
RA Koriati H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,  
RA Tabata S.,  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.  
RT Sequence features of the regions of 1,044,062 bp covered by thirteen  
RT physically assigned P1 clones.";  
RL DNA Res. 4:291-300(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
RA Carinci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamitaya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full length cDNA clones";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
RA Carinci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamitaya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Open Reading Frame (ORF) Clones";  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB0065704; BAB08696.1; -.  
DR EMBL; AY080811; AAL87221.1; -.  
DR EMBL; AY117179; AAM51254.1; -.  
SQ SEQUENCE 806 AA; 92689 MW; 8F8500032F3BCFA2 CRC64;

Query Match 39.7%; Score 1522.5; DB 10; Length 806;  
Best Local Similarity 38.8%; Pred. No. 1.4e-102;  
Matches 304; Conservative 161; Mismatches 223; Indels 95; Gaps 17;

QY 28 EAAAVRALVALPGRPADESVVERALAAKPP-----IDTSLGGGGAARVAVRG 79  
DB 45 QESAKGLRLL-----PTHSQFELRIISKACGCTGCFVLENYGPGRGIGELLIKG 99  
QY 80 STGVAAGAGHRYRDRFCGCVAM---SGSL-RLPSP--LPVAPGE-LTEATPRRYRY 132  
DB 100 TTGVEISGLHMYLKXKCNHVSMDKGGIYAVVPPGHLPRDSKRIFIRRPVNNY 159  
QY 133 QNVCTQSYSEVWDMAREREIDWALNGINALAMSGQEARIQRYVTLALGLTQAEINEF 192  
DB 160 QNVCTQSYSEVWDMAREREIDWALNGINALAMSGQEARIQRYVTLALGLTQAEINEF 219

```

QY 193 FTGPAFLWGRMGNLTHTMDGRLPSPMHIKQLYQHRLVDOKRSFGMTVLPALFAGHVEBA 252
DB 220 FCGPFLAMARNGNLHAMGGRPLSKQMDLQQLQKISRLMKFMTVLVSFSGNVSA 279
QY 253 VTRVPPQVNVYKMGNGHF--NCSYSGFLAPEDDPITPIGSLFLRELKEFG-TDHIY 309
DB 280 LRKIPEANITRLDMNVTVDGSRWCCTYLLNPSDFLEIGEPARKOCTEYGEITRIY 339
QY 310 GADTENMOPSSSEPSYLAATTAYEANTAVTEAAMVLQGMFLQHQPQFMGPAQIRAV 369
DB 340 NQDTNENMPTTSEPEYISLGAAYKAKSKAKAAMLMQGMFLSSDSKTKWPKQLAL 399
QY 370 LCAVPRGRLLVLDLFAESQPVYTRTASFGQOPFIWMLHFGNGHGLFALAEVANGPEBA 429
DB 400 LHSVFPFGKIVLDLVAEVPKPIWKSQAQFGTPYIWMCLHFGNITEMYGAALDSISGVD 459
QY 430 ABLFPGSTWGTGMAPEGISQNEVYYSIMAEIGMRKDPVPDLAAVTSFAARVGVSHPD 489
DB 460 AAVSKNSTWVGWCMCEGIEQNPVYELTSEMAFDEKV-DVQKWLKSYARRRYKENHQ 518
QY 490 AGAARLLLSRYNCSGEACRGNRSPLVRPSLQMTSI----- 529
DB 519 IAAAMELVHTYVNTCT-DGIALHNDPTVKLPDMQSSVQDDLKQKDSWIMSTGPEYTK 577
QY 530 -----WYNSDVFEAMRLLITSAPSLATSPAFRYDLDTROAVQELV 572
DB 578 RRVLPQDKTADLPKALWYSTEKEVIOALKPLEAGDDLSRLTYVYDWVDTLTRYSLDA 637
QY 573 SLVEEARSAYLSKELASLRAGVLA---YELLPALDEVLASDSRFLGSMLEBARAA 629
DB 638 NOYVEAATVAFKQIGSL-----GQISEKFEIJDMDVYLASDNCILGWTLSAKTGA 693
QY 630 VSEADDFEONSRQYLTLMGP-----EGNILDYANKOLAGLVANYTPRMRFLFALVD 684
DB 694 KNGDERKOYEMNARQVTWMTYSDNDVNSKLDYANKFWSGILDEYYLPARALYENEMLK 753
QY 685 SVAQGIPTQOHPQDKNVFQLE---QAFLV-----SKQRPSPQPDGTDVLAKKIFLK 733
DB 754 SFR-----DKKIFKEVEMKREEMIMSHKMQSSSEVYPVAKGDALATISHHLISK 803
QY 734 YYP 736
DB 804 YFP 806

```

## RESULT 7

```

Q92R45 PRELIMINARY; PRT; 811 AA.
AC Q92R45;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE Alpha-N-acetylglucosaminidase.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
CX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bright Yellow 2;
RA Shen W.H.;
RT "Isolation and characterization of a tobacco cDNA encoding a putative
alpha-N-acetylglucosaminidase."
RL Plant Physiol. 118:1534-1534(1998).
DR EMBL; Y18209; CAA77084.1; -
SQ SEQUENCE 811 AA; 93444 MW; E3F07800816996F1 CRG64;

```

Query Match 36.7%; Score 1445; DB 10; Length 811;  
 Best Local Similarity 39.4%; Pred. No. 6.6e-97;  
 Matches 304; Conservative 136; Mismatches 263; Indels 68; Gaps 18;

```

QY 28 EAAAVRALVARLLGPPAADPSVS-VERALAAKPG--LDYSLGGGAARVVRSGTGV 83
DB 45 QESAAGVQLRLL-PAHLSFEFKIYISKOLCGRSCFFRTITNYVSSRNSPBEILIQCTTAV 103
QY 84 AAAAGLHRYLRDTCGGVAM---SSQSL-RLRPP-LPAPVG-ELTEALPNNKRYTQNC 136
DB 104 EITSGHMYLYKRCGAHISMDKTGVQLASVPPGSLPLVKQNELTIQRPVPMNYQNV 163
QY 137 TQSYFVMMNAREREIDWMLNGINLALWAGCAINQRYVALGTLQAEINEFFTG 196
DB 164 TSSYSFVMMQWMEKEIDMTLPGLNPLAFGCGALINQRYVALDYNITTDNDPFGP 223
QY 197 AFLWGRMGNLTHTMDGRLPSPMHIKQLYQHRLVDOKRSFGMTVLPALFAGHVEVTRY 256
DB 224 AFLWGRMGNLTHTMDGRLPSPMHIKQLYQHRLVDOKRSFGMTVLPALFAGHVEVTRY 283
QY 257 FPGVNTYKMGNGHFNCS--YSCSFLAPEDDPFIIGSLFLRELKEFG--TDHIYAD 312
DB 284 FPGANITRLDMNVTVDGSRWCCTYLLNPSDFLEIGEPARKOCTEYGEITRIY 342
QY 313 TENEMOPSSSEPSYLAATTAYEANTAVTEAAMVL-LQGMFLQHQPQFMGPAQIRAVLG 371
DB 343 TENENMPTTSEPEYISLGAAYKAKSKAKAAMLMQGMFLSSDSKTKWPKQLAL 402
QY 372 AVPRGRLLVLDLFAESQPVYTRTASFGQOPFIWMLHFGNGHGLFALAEVANGPEAR 431
DB 403 SVRGRKIVLDLVAEVPKPIWKSQAQFGTPYIWMCLHFGNITEMYGAALDSISGVD 462
QY 432 LFPNSTWGTGMAPEGISQNEVYYSIMAEIGMRKDPVPDLAAVTSFAARVGVSHPDAG 491
DB 463 TSENSITWVGWCMCEGIEQNPVYELTSEMAFDEKV-DVQKWLKSYARRRYKENHQ 521
QY 492 AAMRLLLSRYNCSGEACRGNRSPLVRP-----SLQ 524
DB 522 AAMDILYHTYVNTCT-DGIALHNDPTVKLPDMQSSVQDDLKQKDSWIMSTGPEYTK 580
QY 525 MN-----TSIYNSDVFEAMRLLITSAPSLATSPAFRYDLDTROAVQELV 570
DB 581 WNRRLPFEKSSSLPRLMYSTEDVFQALQFLDKLTKLSGLTYRYDLVDSRSLSK 640
QY 571 LSVLYEARSAYLSKELASLRAGVLA---YELLPALDEVLASDSRFLGSMLEBARAA 629
DB 641 LANQVYLDATSAF-REDDAKFLNQHSKFLPLLQDIDRLAADNFIETWENLCPQNTA 699
QY 630 VSEADDFEONSRQYLTLMGP-----EGNILDYANKOLAGLVANYTPRMRFLFALVD 684
DB 700 WNSDERKOYEMNARQVTWMTYSDNDVNSKLDYANKFWSGILDEYYLPARALYENEMLK 759
QY 685 SVAQGIPTQOHPQDKNVFQLEQAFLVSKORYPSPQPDGTDVLAKKIFLKRY 735
DB 760 SLKXVDFLKEWRKEMIAVSNKMGSTELYPVKAQGDALATATLFEKXF 810

```

## RESULT 8

```

Q95VL5 PRELIMINARY; PRT; 778 AA.
AC Q95VL5; Q95TJ3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE CG13397 protein (LP03571P).
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

```



RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abil J.F., Agbayan A., An H.-J., Andrews-Piankoff C., Baldwin D.,  
 RA Ballew R.M., Baeu P., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Deeson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dushin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Gloder A., Gong F., Gortell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ileguam C.,  
 RA Jatala M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mikhlov G., Mishina N.V., Mobarry C., Morris L., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskey D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA She B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
 RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs S.M., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*,"  
 RL Science 287:2185-2195(2000).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP Celinkner S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Bazon J., An H., Baldwin D., Bazon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K.J., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Honck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ileguam C., Jatala M., Kruse D., Li P., Matel B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svrtkas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
 RT "Sequencing of *Drosophila melanogaster* genome,"  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RN SEQUENCE FROM N.A.  
 RP Mierra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinkner S.E.,  
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richer J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Murgall C.J., Lewis S.E.,  
 RT "Annotation of *Drosophila melanogaster* genome,"  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RN SEQUENCE FROM N.A.  
 RP Adams M.D., Celinkner S.E., Gibbs R.A., Rubin G.M., Venter J.C.,  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RN SEQUENCE FROM N.A.  
 RP FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 [6]

RP SEQUENCE FROM N.A.  
 RC STRAIN-Berkeley;  
 RA Stapleton M., Brokstein P., Hong L., Appayanti A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guan H., Li P., Liao G., Miranda A., Murgall C.J.,  
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celinkner S.,  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003621; AAF52672.2;  
 DR EMBL; AY058738; AAL13967.1;  
 DR FlyBase; FBgn01417; ESTs:172FST  
 SQ SEQUENCE 778 AA; 89746 MW; 9500B1BC4A653B5 CRC64;  
 Query Match 36.4%; Score 1433.5; DB 5; Length 778;  
 Best Local Similarity 41.2%; Pred. No. 4.3e-96;  
 Matches 302; Conservative 120; Mismatches 284; Indels 27; Gaps 13;  
 10 VGVLLAGAGAGAGDEAAAVRALVRLGPGPADPFSYVERALAKGDTYSIGG 69  
 35 MGLQAAHLAPSTPVOVETDAM-AVISRVIGERSQLFKVQVKNM---DLASFQISM 89  
 70 GGAARVVRGSGTGVAAAGLHRYLRDPGCHVANSQSGLRPR--PIPAVGEITATPN 127  
 90 LDDGRILMGMDGVCCKALHLYKLVANKDVDFKRIELPTNLQIPNTVIESKASD- 148  
 128 KRRYQVNTQSYSTVMDMAREREIDWMLNGLNLAASGQEAIMQRYLALGLQA 187  
 149 -IITVQVCTSYSPAMWGIQMRHLDMVLMGLSTLTA-PVQALVWKYTDGLRME 206  
 188 EINEPFTGPAFLAMGRMGLHTWDPPLPSPHIXQLQHRVLDPMRSGMTPLPAPAG 247  
 207 EIDELHAPGAPGWRMNGMNGMGLPFPARRQQLQGEITDQRLGMSVALPARAG 266  
 248 HYPEAVTRFPQVNTKKGSGHFNCSGCSFLAPEDPIPIIGSLFLRLKEFGDH 307  
 267 HVRRLKRLNPESTMEVQRNQPDRYCCGLFEEPTNLKEKIASRLNHTIITYGSH 326  
 308 IYGADTFNMOPSPSESYLAATAVYEAMTAVTEAVMLQGLFQHPQPFMGPQAIR 367  
 327 IFPCPFNELEFPVAKPEYMSATAIYESRIGDIPQALIMQMFYKNF-FWTTDAE 385  
 368 AVLAGVPRGLVLDLPESQPVYTRIASFGQPIFWMLNFGNGLFGALEAVNGP 427  
 386 AFLTPAPGRILVLDLQEGEPFOYELTIRSYGQPIFWMLNFGGLTLMFGSAXLINSGI 445  
 428 EAARLFPSTWGTGMAPEGISONEVYSLMAELGMRDPPDIAAVTSPAARYGVSH 487  
 446 EEARLNSLVGIGITTEGIGQNYMTSFLERQMSNTSL-DLDSWTTNSHRYGYKD 504  
 488 PDGAAMRLILRSVYNSG-EACRGHNSPIVRPSLQMTSINWRSDFEAMRLITS 546  
 505 ERLQAMLLKNSVSPFGLQMRG--QYVTRRPSFQCEPFTWNAVAVDAMILLTF 562  
 547 ASSLATS----PAPFYDLDLTRQAVGELVSLYEEASATLSLALAGVLAELV 602  
 563 RAILEDNRREYEHDDVDITRQFLQISADQLYNLNSARKQVSF---EFLSYKL 618  
 603 LPALDE---VLASDRFLGSLWLEQARAASVSEADPYEONSRYQLTLMWPEGNIIDYA 659  
 619 LKLFDMLILLSSNIFLLGWLQQAQAANTGQQRFFENARNOITAMPGDGIIDYA 678  
 660 NKQAGLVANTTPPWRFLFLALVDSVAGGIPFOHQDQXV-POLEAFVLSKQRYSQ 718  
 679 CQWGLVSDYYRPRWRFLFDVTVALLHAGRFNGTAKLVKYSHELPFSNKQDVYVT 738  
 719 PRGDTVDLAKKIF 731  
 739 PVGNWTLISQDIF 751  
 RESULT 9  
 09AA06 PRELIMINARY; PRT; 770 AA.

AC Q9A06: 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Alpha-N-acetylglucosaminidase.  
 GN CC0540.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;  
 OC Caulobacteraceae; Caulobacter.  
 OC NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; Pubmed=11259647;  
 RA Mierman W.C., Feldblum T.V., Laud M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen U., Heidelberg J.F., Alley M.R.K., Onta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kojanay J.F., Smit J., Craven M.B., Khouri H., Shetty K.,  
 RA Uetreck T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,  
 RA "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AE005727; AAK2527.1; -.  
 DR TIGR; CC0540; -.  
 KW Complete proteome.  
 SQ SEQUENCE 770 AA; 84287 MW; 94A694FA1C18D820 CRC64;  
 Query Match 26.9%; Score 1060.5; DB 16; Length 770;  
 Best Local Similarity 33.0%; Pred. No. 8.1e-69;  
 Matches 260; Conservative 132; Mismatches 302; Indels 93; Gaps 19;  
 QY 2 EAVVAAGVLLAGAGAGDEAREAAVRAIARLILBGPADPSVSEALAKP 61  
 DB 11 QAIISLAAAFAPASALAGSTDG---VAARSLKILPERRLAG-----AHLITTFG 59  
 QY 62 LDT--YSLGGGAARVVRGSGTGVAAAAGLHRYLRDPCGCHVAMSGSLRPLPAPVG 119  
 DB 60 AERGWYAIQKGA-ISISGSPVALVRGAVALRQGLAHVSEGGRRVQAGVAPAG 118  
 QY 120 ETLTATNRYRYQNVCTQSGSYFVWDMARERIDIMMALNGINLALAMSGEALIMORY 119  
 DB 119 ARVE-TDFRHRAYNLTCTGYGTTPTMGMRTRIEDIMMAHAGIDIMPLAMGEQEVWALW 177  
 QY 180 LALILTOELNEFTGPAPFLANGRMGNLHTWDGPLPSPWHIKOLYLORVLDOKRSQMT 239  
 DB 178 REBELSERELADYSGRAFPTMHMGNIIEGYKAPLPTAMIDKDLQVYLIGMRSLGMT 237  
 QY 240 PVLPAFGHVEAVTRYFPQVNVTKMSGWGHFNCSYSCSFLLAEDDIPFIIGSLFREL 299  
 DB 238 PILPAFGYVPAFAEKPKPKARIYRMRPWEGFHETV---WLDPADLPAKIAARFIATY 293  
 QY 300 IKKSGTTHIYAGDLENEMQRP-----SSRPST-----LA 328  
 DB 294 TETAGAGTYIADSFNENLPPINADGADARDAAGDGTANTAVTKKVEYDPAIKQRLA 353  
 QY 329 AATAVYEAMTAVTTEAVMLQGLFOHQPFQWGAQIRAVLGAVERGILLVLDLFAESQ 388  
 DB 354 AYGAIVDSIRQTPDAVWVWQGLFGADSHFWDPALISAVLSLVPDDKMLIIDIGNDY 413  
 QY 389 P-VYTRTASFGQGFPIQMLNFGNGHGLFGAL-----EAVNGGEAARLPENSTWVG 440  
 DB 414 PNWKNAKAPFGGKPMIYGVYVNGGSPVYDGLFPYODIPALIAANDAKL-----AG 467  
 QY 441 TGMAPGISQNEVVYSLAELGMRKDPVPLDAAVTFFAARVYVSPDGAAMRLIRS 500  
 DB 468 FGMPEEGILHNSIYEAAYDLAMSEGAAP-ATWLTYYARARYKTSPLDADALGOVVEA 526  
 QY 501 VYNGSGEACRGHNSP---LVRRPSLQMTSTIY--NRSDVFEAMLLITSAPSLATSP 554  
 DB 527 AFSTRYSPRMWKSAGAYLFFKRPITATVGDPPQHPGRALAEAAVALTALAPTYGOEP 586  
 QY 555 AFRVDLLDTQAOVELSLVYEEARSAVYLSKELASLLRAGGVLAYELLPALEDEVLASDS 614

DB 587 LFLVLDLTATRHILATMKIDDLQVAAVRRGDTA---AGDARVEL-----EXLAISI 637  
 QY 615 RFLIG-----SWEQARAAAVSEAEADFEONSRYOITLWGPENITIDYANKQIAGIV 667  
 DB 638 DKLIGVQDPTLATWIDERARAYGDTPADAAAYAAKKAQVTTIWGGGNLNDYASKWQGLY 697  
 QY 666 ANYTPRRRLPFLBALVDSVAGIPIPGQHOFDKNNFQLEQAFVLSKORY---PSQPRDT 723  
 DB 698 KSPYLPFRMSRFLDAL--KAAGTGFDEVTYTRGVAMERAMVEAVARREKPADPGEI 755  
 QY 724 VDLAKKI 730  
 DB 756 KTLIRI 762  
 RESULT 10  
 ID Q9NAP6 PRELIMINARY; PRT: 715 AA.  
 AC Q9NAP6:  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein K09E4.4.  
 GN K09E4.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Felodetinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smye R.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=990613; Pubmed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018 (1998).  
 DR EMBL; Z83234; CAB70170.2; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 715 AA; 82640 MW; 73CE7FFD05CD995 CRC64;  
 Query Match 26.1%; Score 1027.5; DB 5; Length 715;  
 Best Local Similarity 33.2%; Pred. No. 1.9e-66;  
 Matches 243; Conservative 123; Mismatches 312; Indels 53; Gaps 15;  
 QY 9 AVGVLLAGAGAGDEAREAAVRE-----ALVARLGGPRADPSVSEBALAK 59  
 DB 8 ALLIIVLGA-----ERAGALRVOLQDYDKNSMRSVLRVAFEWLEEFADVKLISAE 62  
 QY 60 PGLDTYSIGGGGAARVRGSG-----TGVAAGLHRYLRDPCGCHVAMSGSLRLP 111  
 DB 63 NG---FQLPKSLAOLIEINKSKRTVTANTPPDALNAINITYRTCLCSQVSNSS--FS 117  
 QY 112 RPLPAVGEILTATENRYRYQNVCTQSGSYFVWDMARERIDIMMALNGINLALAMSG 171  
 DB 118 SCGRKRTSDFINFSKQIRYFNMCITFSYFAWMEPOWERIDWIALNGFTVLMPLG 177  
 QY 172 EALNQRVIALGLTOAELNEFTGPAPFLANGRMGNLHTWDGPLPSPWHIKOLYLORVLD 231  
 DB 178 EILIMDFIGLGVQDELDYSFTSQAYLAMHMGMLKAYGGGLSDAOMLNDHNLAKRIID 237  
 QY 232 QKRSQMTFVLPAFGHVEAVTRYFPQVNVTKMSGWGHFNCSYSCSFLLAEDDIPFI 291  
 DB 238 RLLEIGTPIILPTFGAPFDHLETLFPASKFRRLFRMNFSETSCMLSVSPFDLFPXI 297  
 QY 292 GGLFRELILKEFGTD--HIYADTNEHQPPSS---EPSYLAATTAATYEAATVDTAV 346  
 DB 298 GSTFLRHOKRFGGVVTWNTYSADPNEILPSSAKFAKFAVQOTAOALINSSCKVKNVC 357  
 QY 347 WLLQGLFQHQGFQWGAQIRAVLGAVERGILLVLDLFAESQPVYTRTASFGQGFPIQCM 406

```

Db      358 WVLQSNFFYYDQ--WPMAIKSFLSAIPGNLLILDVAVVPAWQMTSSQGHFFVCL 415
Qy      407 LHNFGNGHGLFGALAEVNGGPEAAALFPGSTWGTGMAPEGISQNEVYVSLAELGMRKD 466
Db      416 LHNFGSRLRLKGLQKIDKGYOLALMKAGSNLVGAGLSMEALDQVYVYQGNIDRMWPE 475
Qy      467 PVPDLAAVTSFAARRVYVSHPDAGAMFLILRSYVNSGEGACRGNHNSPLYRPSLOMN 526
Db      476 PLP-LNNMLKAYSESRSYADPFVQAQKFTLLAGTFEKNQPEKWTGPRFSGVFLYHRGFGSRK 534
Qy      527 TSIWYRSVDFEAMRLLTSS-APSLATSPAFRYDLDLITROAVGELVSLYEEARSAYLS 585
Db      535 IETWPEVEETFSRFRLLPALVHYTGEPHFLFKEDLNDVREMTQ-----FEMGNELALS 588
Qy      586 KEIASLL----RAGVLAAYELLPALEDEVLASDSRFLGSLWEQARAAVSEAEADPYEQN 641
Db      589 MSBAFLMEDKQVGA--SCEMLMEMFQKLESYSNRDVQWIMENAKSIAPTSEERQVFPVT 646
Qy      642 SRQQLTNGPBGNIIDYANKQLAGVANYTYTPRMFLFLALVDSVAQGIPIQGHQFQKNV 701
Db      647 AGDILTWGPTQONLDYAHREWAGLMSGYGRMQTF---CDWILHDDFNHTTFSISV 702
Qy      702 FQ-LEQAFVLS 711
Db      703 FPDVERPFSIS 713

RESULT 11
Q8PPH7 08PPH7 PRELIMINARY; PRT; 798 AA.
AC 08PPH7;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DE N-acetylglucosaminidase.
GN XAC0709.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Camarotte G., Canavan F., Cardoso J., Chambergo F., Clapina L.P.,
RA Clacatelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis U., Wenck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL: AB011700; AAM35598.1; -.
KW Complete proteome.
SQ SEQUENCE 798 AA; EC60943F5A84A7BE CRC64;

```

Query Match 24.4%; Score 962; DB 16; Length 798;  
 Best Local Similarity 32.1%; Pred. No. 1,4e-61;  
 Matches 251; Conservative 123; Mismatches 295; Indels 112; Gaps 23;  
 7 AAAGVLLLAGAGGAAGDEAREAAVRAVLGPGPADPSVSVERALAAKPGIDTYS 66

```

Db      16 AVTLGALLATPVVLA-----ASPAQAVYQRLIG-ARAAQPFMTV---AFRGDADWTR 65
Qy      67 LGGGGAARVRVRSSTGVAAGLHRYLRDFCGHVMSSGQLRPPLPAV-FGELTEAT 125
Db      66 IDAGGT-VRIAGSGVALLARGAVYLIGAGASMSWEGRRVALPAQWPAYRSGV--RT 122
Qy      126 PNRYRYQNTCTQSYSTFWMDPARMERETDMMNLNINLALMSGGEALIMQRYLLGLT 185
Db      123 PPARAVLNTCTGYTTPFMDWPRMQRREIDMMLHGDIMPELAEQGEALIMQALMERFDY 182
Qy      186 CAEINEEFTGPAFLANGRMNLTHTWDGFLPPSWHIRQLYLQHRVLQMSFGMTPLYLPAF 245
Db      183 DDAIAEYFSGRATFPQGMKNITNGYRAPLQCHWIDSKVYLQKILTRMELGNQPLPAF 242
Qy      246 AGHVEAVTRVPQVNTVMKSGWHFNCSYSGFLAPEDPIPIIGSLFLRLIKEFGT 305
Db      243 AGYVPAFAQAHPHAIRRRAWECHETY---WIDPRDPLPAKLARFLLELYAQTYGA 298
Qy      306 DHIYGADTFEMQPP-SSSESYLAAT-----TAVY 335
Db      299 GERFLADATFEMLPVVAADGSDVAAARYGDSIANSDAAKAVPPAQRDARLAEYQALY 358
Qy      336 EAMTAVDTEAVWLLQGLFQHQPFQWPAQIRAVIGAVPRGLVLIDLPAESOP-VYTRT 394
Db      359 RSLAQVVPKATVMQGLFQADRQFWQAQIAAFLGKVPDARLMTVIDIGNDRYFTWKAS 418
Qy      395 ASFGQGFPIWCMLEHNGNGLRG-----ALEAVNGGPEARLFPNSTWGTGMAPEG 447
Db      419 RAFPNGKMITGYVHNTGASNPFLYGDPAFYRHDQLALADBDKXNL-----RFGVFPBG 472
Qy      448 ISQNEVYVSLMAELGMRKDPVPDLAAVTSFAARRVYVSHPDAGAAARLLILRSVNSGSE 507
Db      473 LHSNSVYELVYLAAM-EGHQQSMQWLTHTYLRARYGRSDAALLSAMSDLEAGIYQRYW 531
Qy      508 ACRGHRSP----LVRRPSIQMNTSIWYRSDVF-----AMRLLLTAPSL 550
Db      532 SPKWMNRGAGAYLLFKRPT-----ADIVDDPRDPRDQRLRALDALRQANRY 580
Qy      551 ATPSPAFRYDLD-----LITQAVQELVSLYEEARSAYLSKEIASLLIRAGVLA--YELL 603
Db      581 ADAPLYYDILIEDARHLSLSQADROLQAVY-----QAVDAGDA---RQDAQLAFTQLV 632
Qy      604 PALDEVLASDSRFLGSLWEQARAAVSEAE-ADTFEQNSRYQLTMGPBGNIIDYANKQ 662
Db      633 RGLD-ALVGGQHETLMDWTQAAAAAGHDGLRRAYVGNARAVSWGGGDNLADVASXA 691
Qy      663 LAGLVANYTPRMFLFLALVDSVAQGIPIQGHQFQKNVFOLEQAF-----VLSKQRYPSQ 718
Db      692 WQGYADFYIQRTFRFLSAIRAKKAGTPPDVAIVDHQLATWERSADSDEMSKQPPRE 751
Qy      719 P 719
Db      752 P 752

RESULT 12
Q8XM24 08XM24 PRELIMINARY; PRT; 2104 AA.
AC 08XM24;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE Probable alpha-N-acetylglucosaminidase.
GN CPE0866.
OS Clostridium peifringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,

```

RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
 RT flesh-eater";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
 DR EMBL; AP003188; BAB0572.1; -  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR000421; PA58 C.  
 DR InterPro; IPR003861; FN\_1T.  
 DR Pfam; PF00754; P5\_P8\_type\_C; 1.  
 DR Pfam; PF00041; fn3; 1.  
 DR SMART; SM00606; FN3; 1.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 2104 AA; 236010 MM; 5FE34565959595ED7 CRC64;

Query Match 21.7%; Score 854; DB 16; Length 2104;  
 Best Local Similarity 28.6%; Pred. No. 4.5e-53;  
 Matches 214; Conservative 138; Mismatches 309; Indels 86; Gaps 23;

32 VRLVRLGPGPADPSVERALAAKPGIDTSLGGGGAARVRVSGTGVAAAGLHR 91  
 208 IKNLVGRVIGEEFKDKTFEIRDO--NGNDVEVSDSDRGKVLKNGVSLASGFNY 264  
 92 YLRDFCGCHVAMS--GSQRLRPRPLPAVPGELTEATPNRYRYQVNTOSYSFVWMDMA 148  
 265 YLKQY--CNVSYNPIMSGNLKMPETMPVGERVVIDIPYEXRYALNCTSYTWSFWMD 322  
 149 RWEERIDMALNGINLALMSGQALIMORVYIALGLTOAEINEFTPPAPLARMNLA 208  
 323 QYEEFLDMCANVGLVLDITIGSEVNLRTINEGYDEEKEFISGPAFAFWYQNMNT 382  
 209 TWDGPLPSWHTIKQYLQHRVLIDQMSFGMPTVPAPAGHPEAVTVFPCVNTKMGSV 268  
 383 GFGGLPRMDWEQALRKMDMOSGFINPVIGYSGVPRPFKKNPEAQITSGGM 442  
 269 GHFNCSYSCFLPAB--EDPIFPIIGSLFRELKEFG-TDHIYADTFNEMQPPSS 322  
 443 ---CGFERPMLKTYVNEGEVDYFQVAVDVEYKQKEVFQDVTFNGVDFHE----- 492  
 323 EPSYLAATTA-----VYE---AMTAVDTEAVTWLLQGMFLQGPQFQWPAQIRAVLGA 372  
 493 ---GNTGDDNGKITYELIYNKMTLHNDAYVWLOM-----QGNPSNNKLEGL 538  
 373 VPRGRLVLDLFAESQPYRTTASFOGQPIWCMLEHFGNHLFGALAVNGGPEARL 432  
 539 TKDQAMVLDLFESEVSPDMRLEB-RDLPWIMNMLHFGMGMDAPKEL--ATEIPKA 595  
 433 FPNST--WGTGMAPEGISQNEVYYSIMAELEGKRDPPDLAAVTSFAARRYGVSHDAG 491  
 596 LANEHWVGIGITPEALNTNPLAHELLFDMAWTDQI-NFRTWEDYIERRYGKTKEIL 654  
 492 AAMRLLSVYVNGSEACRGNRSPLYVRPSLQW-NTSIV-----YNSDVFEMRLLL 544  
 655 DAMNILLDTAKKNDYQGAASGLINARPGFGKSMSTWGHKTIYDKSEFEALITFA 714  
 545 TSASPLATSPAPRDLDLTRQAVOEVLSTYEEPARSAVLSKELASLLRAGGVLYALLP 604  
 715 KNYDEFKQSDAFLYDFADILKQILLANSAGEYEWCAVNNNGNEKRFVSGKE-LELIK 773  
 605 ALDEVLTASDSRFLGSMLEQARAASVSEAE--ADFYEONSRYOVLTMGPE-----GNILD 657  
 714 LQERVLSTRPEFLIGNWIEDARTMLKSDSDWTQKLPFENAPALVTTGSRNNADGGGLKD 833  
 658 YANQOLAGIVANVYTPRWLF--LEALVDSVAQGIPOQHDPKNV--FQLEQAFVLSK 712  
 834 YSNQMSGLTEDDYYARWEKWINGLQAEILDGAKA-----FNIDWFMEYDWNKK 884  
 713 ---ORVSPQRGDTVDLAKKIFLKY 735  
 885 SDTDLKLYTEASNEVGLAKIAMEST 911

RESULT 13

Q93TW7  
 ID Q93TW7 PRELIMINARY; PRT; 1862 AA.  
 AC Q93TW7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Mx4E.  
 GN Mx4E.  
 OS Stigmatella aurantiaca.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
 OC Cyctobacterineae; Cyctobacteraceae; Stigmatella.  
 RX NCBI\_TaxId=41;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21110452; PubMed=11182319;  
 RA Sllakowski B., Nordstiek G., Kunze B., Blocker H., Muller R.;  
 RT "Novel features in a combined polypeptide synthase/non-ribosomal  
 RT peptidic synthetase: the myxalimid biosynthetic gene cluster of the  
 RT myxobacterium Stigmatella aurantiaca Sgals5";  
 RL Chem. Biol. 8:59-69(2001).  
 DR EMBL; AF319998; AAK57189.1; -  
 DR InterPro; IPR001227; Ac\_transferrase.  
 DR InterPro; IPR000794; Ketoacyl-synt.  
 DR InterPro; IPR006163; Pp\_bind.  
 DR Pfam; PF00698; Acyl\_transf; 1.  
 DR Pfam; PF00109; ketoacyl-synt; 1.  
 DR Pfam; PF02801; ketoacyl-synt\_C; 1.  
 DR Pfam; PF00550; Pp-binding; 1.  
 DR PROSITE; PS00075; AC\_DOMAIN; 1.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 1.  
 KW Phosphotransferase; Transferrase.  
 SQ SEQUENCE 1862 AA; 200005 MM; E809DCECA320144 CRC64;

Query Match 3.6%; Score 143.5; DB 2; Length 1862;  
 Best Local Similarity 22.3%; Pred. No. 0.35;  
 Matches 180; Conservative 79; Mismatches 289; Indels 255; Gaps 37;

3 AVAAVAVGLLLAGAGAGDEAREAAVRLVRLGPGPADPSVERALAAKPG 62  
 962 AVAPASALVELIGLSAAAEALGAGARELSVEPERALVLEADRRLVQVHLSPASGQHV 1021  
 63 DTYSLGGGGA-----RVVRGSGTGVAAAGLHRLRFGCGCHVAMSQSLRPR- 112  
 1022 HHSRAVGGTSGAGVWRHCRQIRANTPSGSPVLDVAVRRCQHQHGAAYVEQLERC 1081  
 113 ---PLPAB-----PGELTEATPNRYRYQVNTOSYSFVWMDMARERE--IDM 157  
 1082 NVOYEAPLTIGEARRRPGEALGVALSPEVQ-----ESARYQLPALLD-A 1128  
 158 ALNGINLALMSGQALIMORVYIALGLTOAEINEFTGPAFLMG-----RMG 205  
 1129 GLQTLALALAAAGEAV--LFMPPLISGLEVCQ--GRADVKMAHVSIAAATSPEDRMG 1182  
 206 NLTWMDG-----PLPSPWHTIKQVLOH--RVLDQKRS-----FGMTPLPAPAGHVP 250  
 1183 TELDLDBGRRAVAVARVAVARVAVARVAVARVAVARVAVARVAVARVAVARVAVARV 1242  
 251 EAVTVFPQVNTKMGSWGF-----NCSYSCSFLPAE-----DPIPIIG 292  
 1243 WLV-----FLDGGGWGTALVEEIGRQGCCTVAVTAGETFORODARFVVPKPBDM 1294  
 293 SLEFREL-----KRGTDHIYADTFNEMQPPSEPSFLAATAVAYEAM----- 338  
 1295 ERLLELPLALPAGHBRAYVLLGLDAVLDEQGTPESS--VAALHLVAKMSPPARIM 1352  
 339 ---TAVDTEAVWLLQGMFL-----OHQOPWGPAGIRAVLGAVERGLLVLDL 383  
 1353 VYTRGAQVGTQVTERVSLAQAPLMGGSVSLDQGVWGG-----LIDL 1396  
 384 FASQSPYRTTASFOGQPIWCMLEHFGNHLFGALF-----AYNGGPEAR 431  
 1397 APER--VAETAVVIRE-----ISAFGGDGEDQIALREKSLVPRILARGVNAPELR 1448

QY 432 LFNSTMGVGNAPBGISONEVYVYSLMELGNRKDPVDPDLAMVTSFAARRVGVSHPDAG 491  
 DB 1449 LPPDGAIVLTGG-----LGGILGX-----VARMVARGARHL----- 1480  
 QY 492 AAMRLLRBSVNCSGEACRGNHRSPLYVRPSPLOMTSIWYNSDVFEAMRLLITSAPSLA 551  
 DB 1481 ---VLLGRSGASGAGPASAARREGI-----ESLRALGASVTTTLA 1517  
 QY 552 TSPAFRYDLDLTRQAVQELVSLYBEARSAYLSKELASLRAGVLAZEELPALDEVIA 611  
 DB 1518 VVVADEKEXLAALLREAAATL-----PRLRGVIAAALLTESNLNNMD--LA 1561  
 QY 612 SDSRFL-----LGSWL---EQAPAAVSEADPYEONSRQTLTKPBEGLIDYANKOLAG 665  
 DB 1562 AMTAMRPRKVLGSMVLAHEVTR-----EAEIDPFVMSSTS-TLMGASG-LAHYA---AG 1610  
 QY 666 LVANYTTPRMLFLEALV-DVNAQGITP 691  
 DB 1611 -----NOTLEALAHRRAGSLP 1627

## RESULT 14

Q96W93 PRELIMINARY; PRT; 444 AA.  
 AC 096W93;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE TR18.  
 GN TR18.  
 OS *Gibberella zeae* (Fusarium graminearum).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreales; Nectriaceae; Gibberella.  
 OX NCBI\_TaxID=5518;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=88-1;  
 RX MEDLINE=21318682; Pubmed=11425709;  
 RT Lee T., Oh D.W., Kim H.S., Lee J., Kim Y.H., Yun S.H., Lee Y.M.;  
 RT "Identification of Deoxyriboalcohol- and Nivalenol-Producing Chemotypes  
 RT of *Gibberella zeae* by Using PCR.";   
 RL Appl. Environ. Microbiol. 67:2966-2972 (2001).  
 DR EMBL; AF36365; AAKS3574.1; -  
 DR InterPro; IPR005152; LIP.  
 DR Pfam; PF03583; LIP.1.  
 SQ SEQUENCE 444 AA; 47659 MW; BC5208C72484C701 CRC64;

Query Match 3.2%; Score 128; DB 3; Length 444;

Best Local Similarity 20.3%; Pred. No. 0.57;  
 Matches 97; Conservative 54; Mismatches 160; Indels 166; Gaps 24;

QY 157 MALNGIMLALA-MSGGEALWQRYVALGLTQAEINEFFTGPAFLWGRMGNLHTWGPDP 215  
 DB 1 MALDRILFLSLMIG-----FVGAQAALST-----PLP 29  
 QY 216 PS---WHIKOLYLQHRVLDQMSFGMTPLPAPAGHVEAVTRVPOV-NVTMGSGWGHF 271  
 DB 30 PSKXDPMTAP-----PGFENAPGVTLVRPAPGNLTSVTG----- 65  
 QY 272 NCASVSCSFLAPEDPIFP---IGSLFLRELKEFGDHIYADTNEQPSSEPSSTYLA 328  
 DB 66 NCASVYIILYRTTSHKPTWAVTTLLVPELGPESNAHQKIQOSALMSIQVAYDSPDYA 125  
 QY 329 AATTAAYV--EAMTAVDTAVMLQGLFQHPQFQWGA-----QIRAVLG 371  
 DB 126 SPSVTVYTAHISIIYEAA-LAEG-IFVSPDYEGSNAFTAGIISGYATLDSIRAYLS 183  
 QY 372 AVPRGLLVDPRESQPVYTRTASFGQGFIMCMHNGNGLG---ALE-ANNGAP 427  
 DB 184 -----LGLGFNMTNTPVA-----LW-----GYSGALPATEWASELAQVAP 220  
 QY 428 EAALFENSTMGVGNAPBGISONEVYVYSLMELGNRKDPVDPDLAMVTSFAARRVGVSH 487

DB 221 ELVA---GPVGAAGAPLANITSIMIDVNG--GPTSGPTMMLGLT-----VOY 266  
 QY 488 PDAGAAMRLLRBSVNCSGEACRGNHRSPLYVRPSPLOMTSIWYNSDVFEAMRLLITS 547  
 DB 267 PDV-----RKHLVSK-----LNNDSQYNTDFLAAGYITFA 299  
 QY 548 PSLATSPAFRYDLDLTRQAVQELVSLYBEARSAYLSKELASLRAGVLAZEELP 604  
 DB 300 -----GVAFSKIDINK-----YFGKGTIDPSDEPITLVNREGVGYHGP 340

## RESULT 15

Q9FBV4 PRELIMINARY; PRT; 783 AA.  
 AC 09FBV4;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Putative DNA-binding protein.  
 GN SC07844 OR SC00003 OR SC0857.41C.  
 OS *Streptomyces coelicolor*.  
 OC Bacteria; Actinobacteria; Actinomycetidae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Brown S.P., Harris D.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; Pubmed=8843436;  
 RA Redenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmid and a detailed genetic and physical map for  
 RT the 8 Mb *Streptomyces coelicolor* A3(2) chromosome";   
 RL Mol. Microbiol. 21:77-96 (1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=2196410; Pubmed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete *Streptomyces*  
 RT *coelicolor* A3(2).";   
 RL Nature 417:141-147 (2002).  
 DR EMBL; AL939132; CAC03662.1; -  
 DR EMBL; AL939104; CAD30877.1; -  
 DR InterPro; IPR005114; Helicase\_ass.  
 DR Pfam; PF03457; HA; 6.  
 SQ Hypothetical protein; DNA-binding; Complete proteome.  
 SQ SEQUENCE 783 AA; 86432 MW; 545B563958D5EDF CRC64;

Query Match 3.2%; Score 124.5; DB 16; Length 783;

Best Local Similarity 20.4%; Pred. No. 2.4;  
 Matches 141; Conservative 69; Mismatches 249; Indels 233; Gaps 31;

QY 4 VAVAAVGVLLAAGAGACBAAVAVLVRLVILGPGAAVDFSVVERALAAKP--- 60  
 DB 43 LSAAVAGVVRGVWGRGLVIG-EVFLAAGVVRQLASLC-----QVSESHLSRALPSWD 93

```
QY 61 -GLDPTYSIG-----GGGAA-----RVYRGSTGVAAAGLHRYLD-FCGCHVAM 103
Db 94 EGADAKANGEGPNVRFASGSAYPAVMLGCRCTAARTGAACSAFLYSDLSRITCIRHQCW 153
QY 104 SGGSQLRLRPLPAVPELSEA-TPNRYRYQNV-----CTQSYSPV-----WMDW 147
Db 154 SLDSIALDR-VTVVGQVGLAGLPEVIRAHQSLPLLRKSCENAFAPAQAAYVASWWD- 211
QY 148 ARWEREIDMMLNGLNINILAWSGOEAIQRYRYLALGLQAEINEFTGPAFLAWGRMNL 207
Db 212 VHMDEVLMPLRGLGVCDLPBGEVAVLAR-----DVITYPEAVATTVLCD 258
QY 208 HTWDGPLPPSWHIKQLYLQHRVLDQMSFGMTPLYLPAGHVPEAVTRVPEQVAVTKMS 267
Db 259 RLM-----RQVLEBDTH-----GQMPHTLAEV-PRL----- 283
QY 268 MGHFNCYSICSLAPEDPIPIIGSLFLRELKEFGTDHTYGADTFNEMQPSSEPSYL 327
Db 284 -----LTLEARRLGPMVLV-----EQLAASSAGALF 309
QY 328 AAATTAVYEAMTAVDTEAVMLLQGWLFQHQPFQWGPQIRAV-----LGAVRGRLLVLDL 383
Db 310 AMVRACTVRREGAVPEEDVMVAP---VAHRPRGLA-AQVRELRLQHGNAFVGGS- 361
QY 384 PAESQPVYTRTASFOGQPFIMQMLHNFQGNHGLFGALEAVNGPEAPALFPNSTMTVGTGM 443
Db 362 -SRAEQAYRVGLAH-----ASHYAAARHGHLAV-----PKYGR----- 392
QY 444 APEGISQNEVVYSIMAEIGWRKDPVPDLAAWVTSFAARVGV-----HPDA 490
Db 393 -HEGFA-----LGAMV---ANQRTGVALPPIERAQALHRIDPMW 427
QY 491 GAAMFLLLRSYNCGEACRGHNSPLVRPSLQMTSITWNSDVPEAWRLITSAPSL 550
Db 428 NGPMPIISWRRTY-----HRALVHYRKHGLVDATAGFGTSLALGEW----- 468
QY 551 ATSPAFRYDLDLITROAVQELVSLYEBARSAYLSKELASLLRAGVLAAYELPALDEVL 610
Db 469 LHEQCSRYDDLHVGGQRLDLADGIRPAIARSAHPRKSLALAFPAAGLDYARAFAAVHGH 528
QY 611 ASDSR-----FLGSMLEQARA-AVSEAEAD 636
Db 529 ATSKSTRQDGFPLGQWLMGSRBRAMAEKETD 560
```

Search completed: February 13, 2004, 16:17:37  
Job time : 70 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OW protein - protein search, using sw model

Run on: February 13, 2004, 16:14:55, Search time 22 Seconds  
(without alignments)

1428.953 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939  
Sequence: 1 MEAVAVAAAVGVLLAGAG.....VDLAKKIFKKYFGWVAGSW 743

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pdg.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pdg.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pdg.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pdg.\*  
5: /cgn2\_6/ptodata/2/1aa/PTCUS.COMB.pdg.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pdg.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3939	100.0	743	US-09-077-354B-2	Sequence 2, Appl 1
2	123	3.1	3340	US-09-252-991A-23568	Sequence 2, Appl 1
3	121	3.1	4472	US-08-804-227C-2	Sequence 2, Appl 1
4	120	3.0	3729	US-08-804-227C-4	Sequence 2, Appl 1
5	118.5	3.0	1665	US-09-858-664A-2	Sequence 2, Appl 1
6	113.5	2.9	916	US-09-252-991A-23637	Sequence 2, Appl 1
7	113	2.9	7257	US-09-335-409-5	Sequence 5, Appl 1
8	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
9	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
10	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
11	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
12	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
13	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
14	108	2.7	1419	US-09-252-991A-31822	Sequence 5, Appl 1
15	107.5	2.7	405	US-09-252-991A-33024	Sequence 3, Appl 1
16	107.5	2.7	1621	US-08-972-927-2	Sequence 3, Appl 1
17	107.5	2.7	1621	US-08-972-927-2	Sequence 3, Appl 1
18	105	2.7	760	US-09-252-991A-18711	Sequence 6, Appl 1
19	105	2.7	1203	US-09-075-272-4	Sequence 6, Appl 1
20	105	2.7	4150	US-09-428-517-2	Sequence 4, Appl 1
21	104.5	2.7	726	US-08-313-185-49	Sequence 4, Appl 1
22	104.5	2.7	726	US-08-459-499-13	Sequence 13, Appl 1
23	104.5	2.7	726	US-09-083-614A-49	Sequence 4, Appl 1
24	104	2.6	581	US-09-221-235-8	Sequence 8, Appl 1
25	104	2.6	581	US-09-221-928-8	Sequence 8, Appl 1
26	104	2.6	581	US-09-221-527-8	Sequence 8, Appl 1
27	104	2.6	581	US-09-221-236-8	Sequence 8, Appl 1

## ALIGNMENTS

28	104	2.6	581	3	US-09-221-416-8	Sequence 8, Appl 1
29	104	2.6	581	3	US-09-221-245-8	Sequence 8, Appl 1
30	104	2.6	581	3	US-09-163-115-8	Sequence 8, Appl 1
31	104	2.6	581	3	US-09-221-528-8	Sequence 8, Appl 1
32	104	2.6	581	3	US-09-593-553-8	Sequence 8, Appl 1
33	104	2.6	581	3	US-09-221-237-8	Sequence 8, Appl 1
34	104	2.6	668	4	US-09-617-591A-4	Sequence 4, Appl 1
35	104	2.6	677	4	US-08-836-567-2	Sequence 2, Appl 1
36	104	2.6	677	4	US-08-836-567-2	Sequence 2, Appl 1
37	104	2.6	1197	4	US-08-836-567-12	Sequence 12, Appl 1
38	104	2.6	1197	4	US-09-606-304-12	Sequence 12, Appl 1
39	101.5	2.6	2457	4	US-09-252-991A-25671	Sequence 25671, A
40	101	2.6	3519	3	US-09-428-517-4	Sequence 4, Appl 1
41	101	2.6	5215	3	US-09-105-537-2	Sequence 2, Appl 1
42	100.5	2.6	2628	3	US-09-413-814-11	Sequence 11, Appl 1
43	100	2.5	425	4	US-09-252-991A-20467	Sequence 20467, A
44	100	2.5	588	4	US-09-252-991A-31356	Sequence 31356, A
45	100	2.5	1323	1	US-08-026-138E-4	Sequence 4, Appl 1

RESULT 1  
US-09-077-354B-2  
Sequence 2, Application US/09077354B  
Patent No. 625096  
GENERAL INFORMATION:  
APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;  
INVENTOR: WEBER, BIRGIT; BLANCH, LINNE; ANSON, DONALD STEWART  
TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
TITLE OF INVENTION: N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSER: SCULLY, SCOTT MURPHY & PRESSER  
STREET: 400 GARDEN CITY PLAZA  
CITY: GARDEN CITY  
STATE: NEW YORK  
COUNTRY: UNITED STATES  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/077,354B  
FILING DATE: 22-APRIL-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00747  
FILING DATE: 22-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: POKALSKY, ANN R.  
REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 12416  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516 742 4343  
TELEFAX: 516 742 4366  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 743 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 261  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 272  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 435



FEATURE: Potentially-glycosylated Asn site,  
 NAME/KEY: 503  
 LOCATION: 503  
 FEATURE: Potentially-glycosylated Asn site,  
 NAME/KEY: 513  
 LOCATION: 513  
 FEATURE: Potentially-glycosylated Asn site,  
 NAME/KEY: 526  
 LOCATION: 526  
 NAME/KEY: Potentially-glycosylated Asn site,  
 LOCATION: 532  
 US-09-077-354B-2

Query Match 100.0%; Score 3939; DB 3; Length 743;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAAVAAAVGVLLAGAGAGADEARAAARALVRLGPGPADPSVSVERALAAKP 60  
 DB 1 MEAAVAAAVGVLLAGAGAGADEARAAARALVRLGPGPADPSVSVERALAAKP 60  
 QY 61 GIDTVSLGGGGAARVVRGSGTVAAGLHRYLRDFCGCHVAMSGSGLRPLPVPAGE 120  
 DB 61 GIDTVSLGGGGAARVVRGSGTVAAGLHRYLRDFCGCHVAMSGSGLRPLPVPAGE 120  
 QY 121 LREATNRRYVYONCTOSYSPYVMDMAREIDMMLANGINILAMSGCAIMQRYLL 180  
 DB 121 LREATNRRYVYONCTOSYSPYVMDMAREIDMMLANGINILAMSGCAIMQRYLL 180  
 QY 181 ALGLTQAEINEFTGPAFLAMGSMGLHTWDPLPSPWHIKOLYLQHRVLDQMSFGMT 240  
 DB 181 ALGLTQAEINEFTGPAFLAMGSMGLHTWDPLPSPWHIKOLYLQHRVLDQMSFGMT 240  
 QY 241 VLPAPAGHPEAVTRYFPOVNTTKGSGWGHFNCSISCSFLPBPDPPIPIISLRLRLI 300  
 DB 241 VLPAPAGHPEAVTRYFPOVNTTKGSGWGHFNCSISCSFLPBPDPPIPIISLRLRLI 300  
 QY 301 KEFGTHIYGADTFENMOPSPSEPSYLAATAVVEAMTAVDTEAVMLLQGLFQHQPF 360  
 DB 301 KEFGTHIYGADTFENMOPSPSEPSYLAATAVVEAMTAVDTEAVMLLQGLFQHQPF 360  
 QY 361 WGBAQIRAVLGAVERGRLLVLDLFAESQPVYTRTASFOGQPIWCMLEHFGGNHGLFGAL 420  
 DB 361 WGBAQIRAVLGAVERGRLLVLDLFAESQPVYTRTASFOGQPIWCMLEHFGGNHGLFGAL 420  
 QY 421 EAVNGGPEAARLFPNSTVWGTAPEGISQNEVYVSLMAELGMRKOPVDLLAAVTSFA 480  
 DB 421 EAVNGGPEAARLFPNSTVWGTAPEGISQNEVYVSLMAELGMRKOPVDLLAAVTSFA 480  
 QY 481 RRYGVSHPDGAAWRLILRSVYVSCGEACRGNHRSPLVRPSLOMTSITWYRSDVFEAM 540  
 DB 481 RRYGVSHPDGAAWRLILRSVYVSCGEACRGNHRSPLVRPSLOMTSITWYRSDVFEAM 540  
 QY 541 RLLLTAPSLATSPARVYLDLITRQAVGVLSLYEERSAVLSKELASLLRAGVYLA 600  
 DB 541 RLLLTAPSLATSPARVYLDLITRQAVGVLSLYEERSAVLSKELASLLRAGVYLA 600  
 QY 601 ELLPALDEVLASDSRFLGSMLEQAPAAVSEADDFYQNSRYOYLLTGWPGNITLDYAN 660  
 DB 601 ELLPALDEVLASDSRFLGSMLEQAPAAVSEADDFYQNSRYOYLLTGWPGNITLDYAN 660  
 QY 661 KOLAGIVANYYTRRMLFLFALVDSYAGGIPFOQHOFDKNVQLEQAFVLSQKRPSPQR 720  
 DB 661 KOLAGIVANYYTRRMLFLFALVDSYAGGIPFOQHOFDKNVQLEQAFVLSQKRPSPQR 720  
 QY 721 GDTVDLAKKIFLKYYPGWYAGSN 743  
 DB 721 GDTVDLAKKIFLKYYPGWYAGSN 743

RESULT 2  
 US-09-252-991A-23568

Sequence 23568, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 PRIOR FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 23568  
 LENGTH: 3340  
 TYPE: PRT  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-23568

Query Match 3.1%; Score 123; DB 4; Length 3340;  
 Best Local Similarity 21.4%; Pred. No. 0.11;  
 Matches 178; Conservative 101; Mismatches 318; Indels 236; Gaps 44;

QY 4 VAVAA-----AVGVLLAGAGA-----AGDEARAAVRA-LVARTL 40  
 DB 1048 VAIAGERSPOLLYGLAIVAGAGAYVLPDYPSERLAYVLADSGVELLITQAHLEPRL- 1106  
 QY 41 GPG-----PADPSVSVERALAAKPGIDTVSLGGGGAARVVRGSGTVAAGLHRYLRD 95  
 DB 1107 -PGAGVTPICLDSLKLDMNPSQAPGLHLH--GDNLAIVYTSSTGSGQPKVG----- 1156  
 QY 96 FCCGHVAMSGSLRPLPVPAGE--LREATNRRYVYONCTOSYSPYVMDMARE 153  
 DB 1157 --NTHALLA-ERLOMOMATATLADGDDVLMOKAP-----VSFDSVW--E 1195  
 QY 154 IDWMLANGINILAMSGQ-----EAIWQRYLALGLTQAEINEFTGPAFLAMGRMG 205  
 DB 1196 CFMPLVYGCGLVLAAPGHEHDPARIVELVROFVTLHPVPLLOFIDEPFGAAGSILR 1255  
 QY 206 NLHTWDPLPSPWHIKOLYLQHRVLDQMSFGMTVLPAPAGHPEAVTRYFPOVNTTKM 265  
 DB 1256 RLFSGGALPPE-----LRNRYLQ-----LPVALHNRGPTET--AINVT-- 1295  
 QY 266 GSWGHFNCSYCSFLAPED---PIFPIGSLFLELKEFGTHIYGADTFENM--- 317  
 DB 1296 ---HWCC-----RAEDGERSPIGRPLGVNLCVLDLAEF--NILPAGVAGELCIGG 1340  
 QY 318 -----QPSSEPSYLAATAVVEAMTAVDTEAVMLLQGLFQHQPFQHQPF 362  
 DB 1341 LGLARGYLGRPALSAERFVADPLSAAGERLYRTGDRAMNADGVLELGRDQVLRGRF 1400  
 QY 363 ---PAQIRAVLGAVP---RGRLLVLDLFAESQPV--YTRTASFOGQPIWCMLEHFGNH 414  
 DB 1401 RIEPEIICARLLIOPGAOAVVIRREGVAGSOLVGYTTGAVGAEEA-----EQNG 1451  
 QY 415 GLFGALFAVNGGPEAARLFPNSTVWGTAPEGISQNEVYVSLMAELGMRK--DPVPD 470  
 DB 1452 RLRAALQA-----ELPEYVWPAQIMRLAOWPLPS--GKLDTRALPEVWMOQREHVEPRTE 1505  
 QY 471 L-----AAWTSFAARYG-----VSHPDGAAMRLILRSVYVCS-- 505  
 DB 1506 LQRIAVINSEVGLPRVGLRDFELGHSLLATRIVSTRQACVLEPLRLAFASSEL 1565  
 QY 506 -----GEACRGNHRSPLVRPSLOMTSITWYRSDVFEAMRLILTSAPSLATSPAP 566  
 DB 1566 EAPCEQVRAAOAGRTDSHAIIRIDECQVPVPSYSQORWFMFOL-----EPD---SPAY 1618  
 QY 557 RY-----DLDLTR--QAVGELVSLYEERSAVLSKELASLLR--GGV-LAYELL 603  
 DB 1619 NVGGLARLGPDLVARFEALLQALVQ--RHETLRTTPPSVDGVQKRVHGGGLHMDKDF 1677  
 QY 604 PALD-----EVLASDSRFLGSMLEQA---RAAAVSEADDFYQNSRYOYLLTGWPE 652

DB 1678 SALDRSDROCHLQTLA-DSEARHPDLESGLPLRVCMKAEKREHYLVTLHHTIV---TE 1733  
 QY 653 GNLDYANKOLAGVANYTYPRMFLPLEALVDVAQGIPEQOQFQDNVQLE 705  
 DB 1734 GWANDIPARELGALYEAFLEDR-----ESLSELPVO-YLDYSVMORE 1775

# RESULT 3 US-08-804-227C-2

Sequence 2, Application US/08804227C  
 Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.  
 APPLICANT: Kustoss, Stuart A.  
 APPLICANT: Rostock, Paul R., Jr.  
 APPLICANT: Sutton, Kimberly L.  
 TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
 NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501  
 STREET: LILLY CORPORATE CENTER  
 CITY: INDIANAPOLIS

STATE: IN

COUNTRY: USA

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII (DOS) Text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804/227C

FILING DATE: February 21, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 35, 784

REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 4472 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-804-227C-2

Query Match

Best Local Similarity 21.6%; Pred. No. 0.29;

Matches 176; Conservative 71; Mismatches 280; Indels 288; Gaps 44;

DB 4 VAVAAV---GVLILAGAGAGADDEAREAAVAVLVARLLG-----PGPAD 47

DB 1686 VDVAAVVHSGEIAAAVAVAGALSLE--DAAAVVALRAGILGRVLAGAMAAVPLPGE 1743

QY 48 FSVSVERALAAKPGIDTSLGGGAA-----RVVRSGSTVAAAGLH----- 90

DB 1744 -----VEAGLAKPGVEVAAVAVGPASTVSGRRAVAGVAVCAEGVQALLIPVDYASHS 1799

QY 91 RYLDPCG-CHVAVSGSQLRPR-PLPRAV-----PGELTEATPNRRYRYVNTQSTYSFV 143

DB 1800 RHVEDLKKELEKRVLSGIRPSRPVAVCSVAGEOPGEFV----- 1838

QY 144 WMDARMR-----EIDWMLNGINL-----ALAWSGEALIMQRYVLA 181

DB 1839 -FDAGYFRNIRNRYEBAVAVGGLBSGRKRFIVSAHPVIVHLEQTAEAA--DRVYHA 1895

QY 182 LGLTQAE---INEFTGPAFLAMGMENTLWTDGPLPPSMHIKQLYL-----QHRVIDQM 233

DB 1896 TGLTRQDSDPHRLITSTA-EAMAHGATL-TWDPALPFG-HLTTLPYFENHHYMD-- 1950

QY 234 RSGMTVPLPAPAGHVEAVTEVPQVNTKKGSGHFNCSYSGFLLAPEDPIFP-ITG 292  
 DB 1951 -----TTPPTPATTTOSPMDNR-----YRVT-----WK-----ALTESSPVRPHSIG 1988  
 QY 293 SLFRELKEPGTDHVIYGADTFENOMOPSSSEPSYLAATAYEAMTV-----D 342  
 DB 1989 RCLL-----VAPPTDGLDGLTTVISEGASVARELEVIGARR 2028

QY 343 TEAVMLQGMFPOHOPQFMPAQO-----IRAVLGAVPGRGLLVLDLPAESQPYTRTASFO 398

DB 2029 AEVAVEL-----KPSMSAGEENTTVSLGLVP-----STDVARTSIALIQ 2070

QY 399 -----GQPF--IWCL-----HFGGNGLFGALEAVN-----GG-----PEAA 430

DB 2071 AVSDIGVPAARVWALTTRRAVAVPGETPDQAGQLMGFRVALLLPDWGLDLPETA 2130

QY 431 RFPNSTWVGMADEGISQ--NEVYSIMAEIGRKDPVPDLAAMTVSFAARVGVSHP 488

DB 2131 ELTRTPETSQPPOTERLPQTPNRALBELAAVLAGRGEQDQAVRASGIYGRVSRPAA 2190

QY 489 DAGAAMR---LLRSVYNGSEACRGHNSPLVRPSIQNTSITNRSDFEAMRLIL 544

DB 2191 AGAASWQPSGVLITGCGAIGRR-----IARRLAAE-----GAERLVL 2229

QY 545 TS-----APSLATSPAFRYDLD-----LTRQAVOEIVSLY----- 575

DB 2230 TSEKGEAPGAA-----ELAEIRGHCGEVYHACDVAERDALLAALVTPPAVRHT 2282

QY 576 ---YEBARSAYLSKELASILRAGVLAAYELLPALDEVLASDSRFLI-----GSWEQARA 627

DB 2283 AGILDVAVIDLSPSFEVGRKVCAGAEHLQLPADIFVLFSSVGTGWNAGOG 2342

QY 628 A-AVSEADPYEQNSR-----YQLTMGPEG 653

DB 2343 AYAANAALDUALERRRAAGLPATSVAMGLMGGG 2377

## RESULT 4

US-08-804-227C-4

Sequence 4, Application US/08804227C

Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.

APPLICANT: Kustoss, Stuart A.

APPLICANT: Rostock, Paul R., Jr.

APPLICANT: Sutton, Kimberly L.

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501

STREET: LILLY CORPORATE CENTER

CITY: INDIANAPOLIS

STATE: IN

COUNTRY: USA

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII (DOS) Text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804/227C

FILING DATE: February 21, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 35, 784

REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 3729 amino acids

TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-804-227C-4

Query Match  
Best Local Similarity 22.4%; Score 120; DB 2; Length 3729;  
Matches 192; Conservative 72; Mismatches 265; Indels 328; Gaps 47;

4 VAVAAV-----GYLLAGAGAGADBARAAAVRALVRLG-----PGPAD 47  
668 VDVAAVVGSGGIIAAATVAGLSLE--DAAAVLRAELIGRYLAGMAAVLPAGE 725  
48 FSVSVERALAAKGLDLYSLGGGAA----RVVVGSTGVAAAGLH-----90  
726 ---VEAGLAKMGVAAVVAANGPASTVBSGDRRAVAGVAVCAAGVQARLLPYDVASHS 781  
91 RYLDFGCG-CHVAMSGSOLRPR-PLPAV-----PGLTEATPNRYRYQVCTQSFSFV 143  
782 RHVEDLKGELEKRLSGIRPSRPVPCSTVAGEQPEPV-----820  
144 MMDMAER-----EIDMMLNGINL-----ALAMSGQEAIWQRYVLA 181  
821 -FDAGYFNLRRVRFSAVVGGLIEGHRFLEVSAHPVLVHAIEQTAEAA--DRSYHA 877  
182 LGITQAE---INRFFGPAFLAMGEMGNLHTWDGLPSPWHIKOYL-----QHRVLDQM 233  
878 TGLRQODDSPHRLITSTA-EAAVAGATL-TWDPALPFG-FLTPPTPFYFNHHWLD--932  
234 RSPGMPVPLPAF-----AGHVPBAVT--RVFPQVNTYKM--GSMGFNCSYCS 278  
933 ---TPTTPATTTQSPDQAQNPADALPKVSMKRLRDDSLTARLDGM-----978  
279 FLAEPDPIPTIGSLFLAELIK-----ERETD-----IYGADTNNQPPSS 322  
979 LTVVPASADPSVAEGVAEELTAGATVESLTVPEPADSSRLGILLVATERDEAGPIRG 1038  
323 EPEYLA-----AATVAYEAMTAVDTEA-VMLLQGMFLQHPQPFWG 362  
1039 IVSLLALAGDHAGADGAPVVPAGLAASLALIQAGDAGTEAGM-----1083  
363 PAQIRAVLGAVPFGILLVDLFAESQPV---YTRTASFGQPFITCMHNFQGNHGLFG-418  
1084 -AVTRGAVAVAPG-----DVPAPSQALLMGFRVAGIE-LPHCWGLDLDPFGDSGF 1135  
419 ---ALEAVNGGPEARLFPNSTWGT-GMAPEGISONEVYSLMAEL 461  
1136 KQLAATLAGPAEDQVALRA--SGAVGRLLVAASAAGADGWRPG-----TVLV 1184  
462 GWRKDPVPLAAVWTSFAARY---GVSHP---DGAAMRLLSVNGSGEACRGHR 514  
1185 GDTAEVAGPLVWMLNGARRVTLISGLSGPLPEELADVAAVTV-----APCDPADR 1236  
515 SPLVRPSLQMTNISIYNSDVFEAMRLLTSSAPSLATSPARFYDLDTLTRA-----567  
1237 PALRTLLAEPATV-----LVAPPAVPTP-----LAETATALLAIALISA 1277  
568 ---VOELVSLYE-----EARSAYLSKEIASIL 592  
1278 KTGVLVRLDLDLEPPLLEDGLDAFVFPSSVAVGWAGGQGYAAGTAYLD-ALAEGR 1336  
593 RAAG--VLAYELLPADEVLTASRFLGSKLEQA-----RAAY 630  
1337 RAQGLPVTYAAWTFWIG-TPADS---LGEQSKRAGITPLDPAASIDALARAVERRAGCV 1392  
631 SEAEADFYEQNSRYQLT 647  
1393 TVADIDMERFASAYTAT 1409

RESULT 5  
US-09-858-664A-2  
; Sequence 2, Application US/09858664A

Patent No. 6482624  
GENERAL INFORMATION:  
APPLICANT: WEI, Ming-Hui, et al.  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
FILE REFERENCE: C1000927-C1P  
CURRENT APPLICATION NUMBER: US/09/858, 664A  
PRIOR FILING DATE: 2001-05-17  
PRIOR FILING DATE: 2000-11-11  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PASCSEQ for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 1665  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-858-664A-2

Query Match  
Best Local Similarity 18.6%; Score 118.5; DB 4; Length 1665;  
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

19 GGAAGDARAAVRAVRLVRL-LGPGPAD---FSVSVERALAAKGLD-----TY 65  
815 GSASQASSQVSSLRVSSQVTEPSPSLDAEGMTGEADLSDSTPLQRPQCVTRKF 874  
66 SLGS-GGAARVVRGSGTVAAAGLHRYLDFGCG-----CHVAMSGSGLR-----109  
875 SLGRRGGVAVAGVGTAFPGDAG-----GMLGGMMARIMAAVSQSEEEQEARAE 928  
110 ---LPLPAVPELTPNRYRYQVCTQSFSFVMDMAER 151  
929 SQSEEQEARAEELTPVVSARPVPEVGRATRES-----EPFWE 969  
152 ---REI--DMMLNGINLAL---AMSGQEAIWQRYVLA-----182  
970 DIGQSVLQIRDSGDLEADITSLDISEVDPAYLMSLDYIKYLPFERMIRKPKSA 1029  
183 ---GLTQAEINEF-----FTGP-----AFLAMGEMGNLHTWDGP 213  
1030 QPEPSPMAEBELAEFPFTWPMWBGELGPHAGLEITBESDVALLAAAVGKRKWKSSP 1089  
214 LPSPWHIKOYLQHRVLDQNRSGMTPLPAFAGHV-----PEAVTRVFPQVNTYKM 266  
1090 SRSFHPFG---RLPLDEPAELGLRERYKASVEHISRLIKGRPEGLKXGFP---RKKP 1143  
267 SMGFNCSYCSFLLAEPDPIPTIGSLFLAELIKERTDHIYGADTFNMQPPSSPSY 326  
1144 GLASFRLSGLKSWDRAP-----TFLEL-----SDRTV 1172  
327 LAAATVAYEAMTAVDTEAVWLQGMFLQHPQFPGPAQIRAVLGAVERG--RLVLDLF 384  
1173 LGQSVTLACQVSAQPAQATWSKDGALE-----SSRVLISATLKNFOLLTLIVV 1224  
385 AESQPYTRTNSFGQPFITCMHNFQGNHGLFGALEAVNGGPEARLFPNSTWGTGMA 444  
1225 AEDLGVT-----CSVSNALGVTYTTGVLR-----KAER--PSSS-----1257  
445 PEGISQNEVVYSLMAELGWRKDPVPD-----LAAW--VTSFAARYGVSHPDAGAA 493  
1258 ---PDPDIGEYVADGVLLVMKFPVSYGVTVIVQCSLEGGS 1295  
494 WRLLSRYNC-----SGEAC 509  
1296 WTLASDIFDCCVYLSKLSRGVYTFPTACVSKAGMPYSSPSEQVLLGSPSHLASEES 1355  
510 RGHNRSPVVRPSLQMTNISIYNR--SDVFAARLLLTSSAPSLATSPARFYDLDTLR--Q 566  
1356 QGRSAQPLPSTKTAFTQIQRGFSVYRQCEKASRALAAKTIIPHPDKXAVLREYE 1415  
567 AVQELVSLYYEARSAYLS--KEIASILRAGGVLAELLPLADEVLASDSFLLGSKWLEQA 625

Db 1416 ALKGRLPHLAQLHAAYISPHLVLITEL--CGSGELIPCLAE--PASYSESEVVDYIMOM 1472  
 QY 626 AAAA-----VSEADAEYECN--SRVQTLTGMGEINLIDYANKQLAGLVANVTTPMRL 677  
 Db 1473 LSATQYLHNOHLHLDLSENNIITVEYLL-----KVYDIGNNQSLQSEKVLPSDKFD 1566  
 QY 678 FLEALVDSVAQG---IPQCHOFKXNVQL--EQAFVLSKQYPSQPRG 721  
 Db 1527 YLETPAPBLLEGQAVP-----QTDIWAIGVATFIMLSAEYFVSSEG 1568

## RESULT 6

US-09-252-991A-23637  
 / Sequence 23637, Application US/09252991A  
 / Patent No. 6551795  
 / GENERAL INFORMATION:  
 / APPLICANT: Marc J. Rubenfield et al.  
 / TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 / FILE REFERENCE: 107196.116  
 / CURRENT APPLICATION NUMBER: US/09/252,991A  
 / CURRENT FILING DATE: 1999-02-18  
 / PRIOR APPLICATION NUMBER: US 60/074,788  
 / PRIOR FILING DATE: 1998-02-18  
 / PRIOR APPLICATION NUMBER: US 60/094,190  
 / PRIOR FILING DATE: 1998-07-27  
 / NUMBER OF SEQ ID NOS: 33142  
 / SEQ ID NO 23637  
 / LENGTH: 916  
 / TYPE: PRT  
 / ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-23637

Query Match 2.9%; Score 113.5; DB 4; Length 916;  
 Best Local Similarity 21.4%; Pred. No. 0.11;  
 Matches 167; Conservative 77; Mismatches 247; Indels 289; Gaps 42;

QY 1 MEAVVAAAVGVLLLAGAGAA-----GDEAREAAAV--RALVAR-----LL 40  
 Db 238 LQARLAAALGA--GEGAGVAAEQAFQVLRQRRARQRRVPTQGAARHOFLA 295  
 QY 41 GPGPAFDSVVERALAKPGLDTYSLGGGAAR-----VVRGSGTGAATA 87  
 Db 296 GTGLAVDQRRVERRHARTGLQ-----GADRLGIAEHLBPLGVMVQGRQALDAV 348  
 QY 88 GLHRLRDFCGCHVAMSGSO-----LRLPRPLAVPEGLTEAPNRYRYOVNCTQSYS 141  
 Db 349 ---RFV-----EGQAAGVPLALANPFQDRGIEQGLAHGHPQR----- 388  
 QY 142 FVWMDMAREREIDMALNGINLALAMSGORAIWQRYVALGLTQAEINEFTGPAPLAW 201  
 Db 389 ---CAEALDQRIIGFEIVEQA--GQPLAGR-----VASALAE 422  
 QY 202 GRMGTLHTWDGRLPSPWHIKOLYLQHRVLDQMSFGMTFVLPAPAGHVPEAVTRVFQVN 261  
 Db 423 GRVGQHR-----QALAVHROHRI-----GHRQQRLELQAAAL 455  
 QY 262 VTKMGSWGH---FVCSYSCFLLAPEDPIPIIGSLFLRELIKEFGTD----- 306  
 Db 456 AGKVDHHRHLATDAQORCA-----QFLGHFEGEGRVVDVVRH 496  
 QY 307 ---HIVG--ADTFENEMQPPSSBSYIAAATTVYEMTAVDTEAVML 349  
 Db 497 HFRHIOVVAFTQORHLLGDADTVDEGDVAHAGTW--TAATLAIKISIVAPGA-- 550  
 QY 350 QGMFLFOHOPFMGPAQ---IRAVLGAVRGRLLVLDLFAESOPV--YTRTASFGQGFIV 404  
 Db 551 ---QPK--GRPSGSGAQLAAGAVAPALVHHGFGHQRARLLRVSLVQ----- 595  
 QY 405 CMLHNFSGNHGLFGALFAVNGGPEARLLFPNSTVY-----GTGAAPBSIGSENVYS 456  
 Db 596 ---HDFG-----LGIVPFGGD--VRQHPHPLAVADPQQRRLPGAGPPTQQLALAA 643

QY 457 LAEE-----LGRKDP--VPDLAMVTSFAAR-----RYG 484  
 Db 644 AVEAEADEBHRGATAGRLG--RDARVGAPEVAALIFVVAQCGAVRRVGHYPGEVRG 701  
 QY 485 VSHPDGAAMRLLLRSEVNSGCBACGHNRSPULVRPSL--QMTNSTWNRSDVPEAWML 542  
 Db 702 TG--EHAASQGVAR--VDGAVGDLVALADBAFGIARPGALGQBPRTVAQS--VEAY-- 754  
 QY 543 LITSAPSLATSPAFRYD-----LLDT-----TRQAVOEIVSLYEAREAGAYLS 585  
 Db 755 LIDEALHVAFAVARRHDLRRSDALLEVGMAGBEGLLGGEGALRLARLAIIVQARRIDPF 814  
 QY 586 KEIASILRAGVLAAYLLPALDEVLASDSRFLLSGLTWEPARAASAEADAEYEQNSRQ 645  
 Db 815 QGLAALPVGG-----LPLAQ--AED--FAFGSLDQ--RVQAEERGAERYAESARH 863

## RESULT 7

US-09-335-409-5  
 / Sequence 5, Application US/09335409  
 / Patent No. 6121029  
 / GENERAL INFORMATION:  
 / APPLICANT: Schupp, Thomas  
 / APPLICANT: Ligon, James  
 / APPLICANT: Molnar, Istvan  
 / APPLICANT: Zirkle, Ross  
 / APPLICANT: Cyr, Devon  
 / APPLICANT: Goetlich, Joern  
 / TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLIONES  
 / FILE REFERENCE: 4-30582A  
 / CURRENT APPLICATION NUMBER: US/09/335,409  
 / CURRENT FILING DATE: 1999-06-17  
 / NUMBER OF SEQ ID NOS: 30  
 / SOFTWARE: PatentIn Ver. 2.0  
 / SEQ ID NO 5  
 / LENGTH: 7257  
 / TYPE: PRT  
 / ORGANISM: Sorangium cellulosum  
 US-09-335-409-5

Query Match 2.9%; Score 113; DB 3; Length 7257;  
 Best Local Similarity 22.3%; Pred. No. 4;  
 Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPSWHIKOLYLQHRVLDQMSFGMTFVLPAPAGHVPEAVTRVFQVNTKX--GSGWHN 272  
 Db 4805 IPAEWPLQGVITHAAGALD-----GVLDQTTDRFGRVLA PKVTGAW----- 4846  
 QY 273 CSYSCSFLLAPEDPIPIIGSLFLRELIKEFGTDHIYGADTFENEMQPPSSBSYIAAAT 332  
 Db 4847 ---NHLELTAGNDLAFVFLFS--SMGGLLSAQSQSYAANFTLDLAAHRRAREGLAASL 4902  
 QY 333 A-----VEAMRAVDTEAVMLQGMFLFOHQPFGWPAQIRAVLG--AVPRGL--L 379  
 Db 4903 AMGPWSDGMAAGLSA-----LQARLARHGMGLSPAQGTALLQALAPFETQLGAM 4955  
 QY 380 VIDLFAESQPVYTRTASFGQGFIVCMLNPFQGNHGLFGALFAVNGGPEARLLFPNSTV 439  
 Db 4956 SLDPRAAQ-----ASGAAPFVWRALVRABARTAGAACAL-----AARL 4997  
 QY 440 GTGMAPEGISQNEVYISLAE-----LGR-----KQVPPL-----AAVTSFAAR 482  
 Db 4998 --GALPEARRADEVRKVOAELARVLSMSASAVPDRPLSLDGLDLSLTAVELRVNLGQR 5055  
 QY 483 YGVG-----HPDGAAMRLLLRVYNGSGEACRGNHRSPLV-----RRP 521  
 Db 5056 VGATLTPATLAFHPVTDALTRWLDKVLVAAPSVSAKSPQVALDERIALIIGICRPP 5115  
 QY 522 SLQMTNSTWNRSDVPEARLLITSAPSLATSPARRYLLDITRAVQDELVLYYE--E 578  
 Db 5116 G-----GVADEPSFWRLLBEGSDAVEVEHRRWD-----IDAFYDPDP 5154  
 QY 579 ARSAYLSKELASLBRAGVLA-----YELLPALDEVLASDSRFLI--GSMLEQARAA 628





```

; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-568-472-5
Query Match      2.9%; Score 113; DB 4; Length 7257;
Best Local Similarity 22.3%; Pred. No. 4;
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSMHIKQLYGHRVLDQKRSFGMTFVLPAFAGHVEATRVPPQVNTKMGSMGHFN 272
DB IPAEWPIQGVYHAAGLDD-----GVLEQCTTDFRSYLAQKVTGAM----- 4846
QY 273 CSYSCSFLAPEDPIPIIGSLFLRELKEFGTDHYGADTFNEMQPPSEPSYLAAT 332
DB 4847 ---NLHETLGNLAFVLFSSMSGLGSAQGSNYAANTFLDLAAHRAAGLAAQSL 4902
QY 333 A-----YEAMTAVDTFAVWLQGMLEFOHOPQFPGAOIRAVIG---AVPRGRL--L 379
DB 4903 AMGPWSDGMAAGISAA-----LQRLARHGMGALSPAOGTALGQALARPETQIGAM 4955
QY 380 VLDFEASQPYRTTASFOGQPIWCMLEHFGNGHGLFGALEAVNGGPEARLFPNSTW 439
DB 4956 SLDVRAASQ-----ASGAAPFPWRALVBAERHTAAGQAL-----AARL----- 4997
QY 440 GTGMAPEGISONEVYSLMAE---LGWR-----KDPVPDL-----AAVTSFAAR 482
DB 4998 --GALPEARADEVRKVKQAEIARVLSWSAASAVPVDRLSDGLSLTAVELRVLGQR 5055
QY 483 YGVS-----HPDGAARLLRSYVNCSGACRGNRSPV-----RRP 521
DB 5056 VGATLPATLAFDHTYDALTRWLDKVLAAEPVSASAKSPVALDEPIAIGICRFP 5115
QY 522 SLQNTSITWNRSDVEAFAMRLTSPATSPAFRYDLIDLTRQAVQELVSLYE---E 578
DB 5116 G-----GVADPESFWRLLEEGSDAVEVHERMD-----IDAFYDPPD 5154
QY 579 ARSAYLSKELASLRAGVLA-----YELLPALDEVLASDRFLL--GSMLEQARAA 628
DB 5155 VRGKMTT-----RFGGLSDIDRFPDPAFGISPREATTMPQORLLLETSMWAFERAG 5207
QY 629 AVSE 632
DB 5208 ILPE 5211

RESULT 13
US-09-567-899-5
; Sequence 5, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Lagon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-567-899-5
Query Match      2.9%; Score 113; DB 4; Length 7257;
Best Local Similarity 22.3%; Pred. No. 4;
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

```

```

Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSMHIKQLYGHRVLDQKRSFGMTFVLPAFAGHVEATRVPPQVNTKMGSMGHFN 272
DB IPAEWPIQGVYHAAGLDD-----GVLEQCTTDFRSYLAQKVTGAM----- 4846
QY 273 CSYSCSFLAPEDPIPIIGSLFLRELKEFGTDHYGADTFNEMQPPSEPSYLAAT 332
DB 4847 ---NLHETLGNLAFVLFSSMSGLGSAQGSNYAANTFLDLAAHRAAGLAAQSL 4902
QY 333 A-----YEAMTAVDTFAVWLQGMLEFOHOPQFPGAOIRAVIG---AVPRGRL--L 379
DB 4903 AMGPWSDGMAAGISAA-----LQRLARHGMGALSPAOGTALGQALARPETQIGAM 4955
QY 380 VLDFEASQPYRTTASFOGQPIWCMLEHFGNGHGLFGALEAVNGGPEARLFPNSTW 439
DB 4956 SLDVRAASQ-----ASGAAPFPWRALVBAERHTAAGQAL-----AARL----- 4997
QY 440 GTGMAPEGISONEVYSLMAE---LGWR-----KDPVPDL-----AAVTSFAAR 482
DB 4998 --GALPEARADEVRKVKQAEIARVLSWSAASAVPVDRLSDGLSLTAVELRVLGQR 5055
QY 483 YGVS-----HPDGAARLLRSYVNCSGACRGNRSPV-----RRP 521
DB 5056 VGATLPATLAFDHTYDALTRWLDKVLAAEPVSASAKSPVALDEPIAIGICRFP 5115
QY 522 SLQNTSITWNRSDVEAFAMRLTSPATSPAFRYDLIDLTRQAVQELVSLYE---E 578
DB 5116 G-----GVADPESFWRLLEEGSDAVEVHERMD-----IDAFYDPPD 5154
QY 579 ARSAYLSKELASLRAGVLA-----YELLPALDEVLASDRFLL--GSMLEQARAA 628
DB 5155 VRGKMTT-----RFGGLSDIDRFPDPAFGISPREATTMPQORLLLETSMWAFERAG 5207
QY 629 AVSE 632
DB 5208 ILPE 5211

RESULT 14
US-09-252-991A-31822
; Sequence 31822, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31822
; LENGTH: 1419
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-31822
Query Match      2.7%; Score 108; DB 4; Length 1419;
Best Local Similarity 21.9%; Pred. No. 0.76; Indels 168; Gaps 23;
Matches 100; Conservative 40; Mismatches 148;

QY 349 LOGMLFOHOPQFPGAOIRAVIGAVP-----RGRLVLDFEASQPYRTTA---SFOG 399
DB 178 LAGMARBAAGVLRG---LRIGIGARPGEWMDGNNLIDTPRTDAQVVEALGERLQG 234
QY 400 --OPTWCMLEHFGNGHGLFGALEAVNG-----PEARLTPNSNMGTGMAPE 446
DB 235 GTQPMI-----RVGGGPPFGNLPALGGLVGVYANGGANLRLLFEAVGRWHAQLPVD 289

```



```
QY 447 GISONVVYSIMAEIGWRKDPV-----DLAAWTS-----FAARRYGVSHPD 489
      : : : : :
Db 290 ALPAPQP-----LAQAGFHPDAPAFAGIADYLAWGASRWASDAPRIAFIIPGALADAQ 345
      : : : : :
QY 490 AGAAMRLIRSVYNSGACGHRSPVRRPSPLOMTSTIYNRSDVFEAMR----- 541
      : : : : :
Db 346 TGAIDELIRRS-----ERHGOAPL-----AVWFDSDP-EALRKSFAQADV 385
      : : : : :
QY 542 LLTSPASPLATSPAERYDLDTROAVQELVSLYEE-----ARSAYLSKELASL-- 591
      : : : : :
Db 386 QALVVLQHLONGPARAFALD--VPVLQTLGRDGNENDWLAASGVAFRTVAFLG 442
      : : : : :
QY 592 -----LRAGVLAYPELPAIDEVL-----ASDSRFLGSMV- 622
      : : : : :
Db 443 MPETWMSDPLVISALENGEPELMAG--QAEALDKLDRLRLRPVADKHALMFMNH 500
      : : : : :
QY 623 -----EQRAAAVSEAEADFEQNSRYQLTLMGPEGNILDYAN 660
      : : : : :
Db 501 PEGENVAASHLVNVPASLARGELRAAGYRVATSD-----ESALIDTAQ 545
      : : : : :
QY 661 KQLAGLVNYYTPRRRLFLBALV-DSVAGGIPFOQH 695
      : : : : :
Db 546 RLGG---YYRPQ---TLDALYRDGLAASLPIDAY 574
      : : : : :
```

## RESULT 15

```
US-09-252-991A-33024
; Sequence 33024, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107136.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 33024
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-33024
```

Query Match 2.7%; Score 107.5; DB 4; Length 405;

Best Local Similarity 35.6%; Pred. No. 0.1; 57; Indels 19; Gaps 8;

Matches 47; Conservative 9; Mismatches 9;

```
QY 8 AAVGVLLAGA---GAA---GDEAREAAVRLVATLGGPAPADFSVSEBALAKP 60
      : : : : :
Db 251 AAGGAVALLRGCMRGRGAORPGPRRRPAPAGM--RAVGKPVAGVA---RRLAGEP 305
      : : : : :
QY 61 GID--TYSLGGGGAARVVRGS-TGVAAAAGLHRYLRDFCGCHVAMSGSLRPLPAY 117
      : : : : :
Db 306 VROPAPALGGGAGARALVGGELAGIAGATRHAGRS--GTHRRPRA--GLPQASDV 361
      : : : : :
QY 118 PGLTEATPNRY 129
      : : : : :
Db 362 PGLVFRFRRA 373
      : : : : :
```

Search completed: February 13, 2004, 16:18:02  
Job time : 26 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2004, 16:15:30 / Search time 56 Seconds  
(without alignments)  
2778.054 Million cell updates/sec

Title: US-09-836-613-2  
Perfect score: 3939  
Sequence: 1 MEAAVAAGVAVLLAGAGC.....VDLAKKIFLKYPGWVAGSW 743

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3939	100.0	743	11	US-09-836-613-2
2	2120	53.8	402	12	US-10-029-386-33001
3	832.5	21.1	728	15	US-10-156-761-9552
4	827.5	21.0	1038	15	US-10-156-761-13515
5	147.5	3.7	101	12	US-09-864-408A-3432
6	121	3.1	4342	9	US-09-815-242-5107
7	118.5	3.0	469	15	US-10-156-761-9695
8	118.5	3.0	871	15	US-10-367-019-7
9	118.5	3.0	1351	15	US-10-367-019-1
10	118.5	3.0	1610	15	US-10-367-019-4
11	118.5	3.0	1665	9	US-09-858-664A-2
12	118.5	3.0	2596	15	US-10-307-019-6
13	118.5	3.0	2630	14	US-10-077-130-2
14	118.5	3.0	7968	14	US-10-077-130-5
15	117.5	3.0	497	12	US-09-820-788-4

16	117.5	3.0	497	12	US-10-313-963A-60	Sequence 60, Appli
17	117.5	3.0	767	15	US-10-156-761-7798	Sequence 7798, Ap
18	115.5	2.9	3668	12	US-10-402-842-4	Sequence 4, Appli
19	114	2.9	473	12	US-10-368-493-10510	Sequence 10510, A
20	114	2.9	823	12	US-10-368-493-10184	Sequence 10184, A
21	113.5	2.9	567	12	US-10-368-493-16185	Sequence 16185, A
22	113.5	2.9	619	12	US-10-368-493-15803	Sequence 15803, A
23	113.5	2.9	771	12	US-10-368-493-15435	Sequence 15435, A
24	113	2.9	717	12	US-10-368-493-17769	Sequence 17769, Ap
25	113	2.9	7257	14	US-10-014-717-5	Sequence 5, Appli
26	109.5	2.8	1377	9	US-09-815-242-10384	Sequence 10384, A
27	109.5	2.8	1377	12	US-10-287-274-567	Sequence 467, App
28	108.5	2.8	796	15	US-10-156-761-14345	Sequence 14345, A
29	106.5	2.7	4498	15	US-10-156-761-15905	Sequence 1905, Ap
30	106	2.7	308	10	US-09-738-626-5728	Sequence 6728, Ap
31	106	2.7	310	9	US-09-826-909-2	Sequence 2, Appli
32	105.5	2.7	527	12	US-10-368-493-10548	Sequence 10548, A
33	105	2.7	4150	12	US-09-808-880-2	Sequence 2, Appli
34	104.5	2.7	726	11	US-10-368-493-23643	Sequence 23643, A
35	104	2.6	581	9	US-09-799-875-11	Sequence 11, Appli
36	104	2.6	581	9	US-09-757-982-8	Sequence 8, Appli
37	104	2.6	654	10	US-09-909-320-177	Sequence 177, App
38	104	2.6	654	10	US-09-909-0888-177	Sequence 177, App
39	104	2.6	654	10	US-09-905-281A-177	Sequence 177, App
40	104	2.6	654	10	US-09-902-883-177	Sequence 177, App
41	104	2.6	654	10	US-09-907-824-177	Sequence 177, App
42	104	2.6	654	10	US-09-907-841-177	Sequence 177, App
43	104	2.6	654	11	US-09-904-011-177	Sequence 177, App
44	104	2.6	654	11	US-09-906-742-177	Sequence 177, App
45	104	2.6	654	11	US-09-906-838-177	Sequence 177, App

## ALIGNMENTS

RESULT 1  
US-09-836-613-2  
Sequence 2, Application US/09836613  
Publication No. US20030039643A1  
GENERAL INFORMATION:  
APPLICANT: HOBWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;  
WEBER, BIRGIT; BLANCH, LINNIE; ANSON, DONALD STEWART  
TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES  
ENCODING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: NIXON PEARODY LLP  
STREET: 990 STEWART AVENUE  
CITY: GARDEN CITY  
STATE: NEW YORK  
COUNTRY: UNITED STATES  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/836,613  
FILING DATE: 17-Apr-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00747  
FILING DATE: 22-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: POKALSKY, ANN R.  
REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 2249/104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516 742 4343  
TELEFAX: 516 742 4366  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 743 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 261  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 272  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 435  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 503  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 513  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 526  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 532  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-836-613-2

Query Match 100.0%; Score 3939; DB 11; Length 743;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAVAVAAAVGVLLGAGAGAGDEARBAARALVARLIGPPADPFVSVERALAAAP 60  
DB 1 MEAVAVAAAVGVLLGAGAGAGDEARBAARALVARLIGPPADPFVSVERALAAAP 60  
QY 61 GUDTYSIGGGGAARVRVRSSTGVAAAAGLHRYLRDFCGCHVAVSSQQLPPLPAVGE 120  
DB 61 GUDTYSIGGGGAARVRVRSSTGVAAAAGLHRYLRDFCGCHVAVSSQQLPPLPAVGE 120  
QY 121 LTRATPRRYRYNVCOSYFVWMDMARWEREDIMMANGINLALAMSGOARIMQRYL 180  
DB 121 LTRATPRRYRYNVCOSYFVWMDMARWEREDIMMANGINLALAMSGOARIMQRYL 180  
QY 181 ALGLTOAINEFTGPAFLAMGRGNLHTWGPJPSWHIKQYLRVLDQMSFGMT 240  
DB 181 ALGLTOAINEFTGPAFLAMGRGNLHTWGPJPSWHIKQYLRVLDQMSFGMT 240  
QY 241 VLPAPAGHVEATRVPRQVNVTKMGSGHFNCSYCSFLAPEDIFPLISLFLREI 300  
DB 241 VLPAPAGHVEATRVPRQVNVTKMGSGHFNCSYCSFLAPEDIFPLISLFLREI 300  
QY 301 KEFGTDHIGADTFENEMQPPSSPSYLAATTAVYAMTAVDTAVVLLQGMFQHPQF 360  
DB 301 KEFGTDHIGADTFENEMQPPSSPSYLAATTAVYAMTAVDTAVVLLQGMFQHPQF 360  
QY 361 WGAPOIAAVYGAIVPRGRLVLDLFAESQPYRTTASFGQOPRTMCLAHFGNGHGLFGL 420  
DB 361 WGAPOIAAVYGAIVPRGRLVLDLFAESQPYRTTASFGQOPRTMCLAHFGNGHGLFGL 420  
QY 421 EAVVAGGEAARLFPNSTVGTGMAPEGISONEVYVSLMAELGWRKDPVDLAAWTSFAA 480  
DB 421 EAVVAGGEAARLFPNSTVGTGMAPEGISONEVYVSLMAELGWRKDPVDLAAWTSFAA 480  
QY 481 RRYGVSHPDGAARLLRSVYVNCSGACGHRSPVLRPSIQNTSTMYNRSDVFEAM 540  
DB 481 RRYGVSHPDGAARLLRSVYVNCSGACGHRSPVLRPSIQNTSTMYNRSDVFEAM 540  
QY 541 RLILTSAPSLATSPAFRYDLDLTRQAVQELVSLYEEARSAVLSKELASILFAGVLA 600  
DB 541 RLILTSAPSLATSPAFRYDLDLTRQAVQELVSLYEEARSAVLSKELASILFAGVLA 600  
QY 601 ELIPALDEVLASDSRFLGSMLEQARAASVSEADPFYEQNSRYQLTWGPBGNILIDYAN 660

DB 601 ELIPALDEVLASDSRFLGSMLEQARAASVSEADPFYEQNSRYQLTWGPBGNILIDYAN 660  
QY 661 KQLAGLVANYTTPMRFLFBLAVDSVAQGIPOOHQPKVNFQLEQAFVLSKORYPSPQR 720  
DB 661 KQLAGLVANYTTPMRFLFBLAVDSVAQGIPOOHQPKVNFQLEQAFVLSKORYPSPQR 720  
QY 721 GDTVDLAKKIFLKYPGAVAGSW 743  
DB 721 GDTVDLAKKIFLKYPGAVAGSW 743

RESULT 2  
US-10-029-386-33001  
Sequence 33001, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Hanzel, David R.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: AEMTCA-X-2  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 3428  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 33001  
LENGTH: 402  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO U34879.1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.74  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.93  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6  
OTHER INFORMATION: SWISSPROT HIT: P54802, EVALUO 0.00e+00  
US-10-029-386-33001

Query Match 53.8%; Score 2120; DB 12; Length 402;  
Best Local Similarity 100.0%; Pred. No. 6; 9e-185;  
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 DTEAVMLIQGMFQHPQFPGPAQIRAVLGAIVPRGRLVLDLFAESQPYRTTASFGQOP 401  
DB 1 DTEAVMLIQGMFQHPQFPGPAQIRAVLGAIVPRGRLVLDLFAESQPYRTTASFGQOP 401  
QY 402 FIKMCLAHFGNGHGLFGLLEAVNGGPEAARLFPNSTVGTGMAPEGISONEVYVSLMAEL 461  
DB 61 FIKMCLAHFGNGHGLFGLLEAVNGGPEAARLFPNSTVGTGMAPEGISONEVYVSLMAEL 461  
QY 462 GWRKDPVDLAAWTSFAARVGVSHPDGAARLLRSVYVNCSGACGHRNSPLVRP 521  
DB 462 GWRKDPVDLAAWTSFAARVGVSHPDGAARLLRSVYVNCSGACGHRNSPLVRP 521  
QY 521 SLOMNTSTMYNRSDVFEAMRLLTSAPSLATSPAFRYDLDLTRQAVQELVSLYEEARS 581  
DB 121 GWRKDPVDLAAWTSFAARVGVSHPDGAARLLRSVYVNCSGACGHRNSPLVRP 180  
QY 582 SLOMNTSTMYNRSDVFEAMRLLTSAPSLATSPAFRYDLDLTRQAVQELVSLYEEARS 581  
DB 181 SLOMNTSTMYNRSDVFEAMRLLTSAPSLATSPAFRYDLDLTRQAVQELVSLYEEARS 581  
QY 641 AYLSKELASILFAGVLAAYELLPALDEVLASDSRFLGSMLEQARAASVSEADPFYEQN 641  
DB 241 AYLSKELASILFAGVLAAYELLPALDEVLASDSRFLGSMLEQARAASVSEADPFYEQN 641  
QY 642 SRYQLTWGPBGNILIDYANQKLAGLVANYTTPMRFLFBLAVDSVAQGIPOOHQPKVNF 701  
DB 301 SRYQLTWGPBGNILIDYANQKLAGLVANYTTPMRFLFBLAVDSVAQGIPOOHQPKVNF 701  
QY 702 FQLEQAFVLSKORYPSPQRPDVTDLAKKIFLKYPGAVAGSW 743  
DB 361 FQLEQAFVLSKORYPSPQRPDVTDLAKKIFLKYPGAVAGSW 743

## RESULT 3

US-10-156-761-9552  
 ; Sequence 9552, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; PRIOR FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 9552  
 ; LENGTH: 728  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-9552

Query Match 21.1%, Score 832.5; DB 15; Length 728;  
 Best Local Similarity 31.9%; Pred. No. 8.3e-67;  
 Matches 239; Conservative 106; Mismatches 338; Indels 67; Gaps 21;

12 VLLAGAGGAGDAREAAVRLVATLGGPDAADSVSVERALAAK-----PG 61  
 7 VLLTLLAAGSG-----AAVACDPASADG-APAGSAAATARRLLPRHRQLTFRAGG 59  
 62 LDYSLGGGGAARVRVSGTGVAAAGLHRYLDFCGCHVAMSGSGLRPRPLPAVAGEL 121  
 60 RDTFOV-SGRTGRVTVGGTATQTLGLNMYLRNIADINMAGROLRLPRALPGLAGTV 118  
 122 TEATPNRYRYQNVCTQSYFVWMDMAREREIDMMALNGINILAMSGOEALMORVYLA 181  
 119 TRRANVRERFALNTNDGTYTPYHDWTYMERELDLVLAHGNEVLVGTGADLHHRVFOE 178  
 182 LGLTQAEINEFTGPAFLAMGRMGNLHTWDGRLPPSWHIKOLYLOHRVLDQMSFGMTV 241  
 179 FGTYDELRKMIPEGALGPWWLLQNLGAFRPVSGQLDARALGRRIARLRLBELMTV 238  
 242 LPFAGHVPBAVTVFPQVNVTKMGSGHFNCSYSCSFLAPEDPIPTIGSLFRELIX 301  
 239 FPGYFGTVPPGAFDRNAGAHVPOGTW---MGFARPDWLDPRTEHFRVAAAFYRQDE 294  
 302 EF-GTDDHYGADTENNOMPPSEPSYLAATTAAYEAMTAVDTEAVVLLQGMFOHOPF 360  
 295 MFGASSTRYKMDLHEGSGPDVP--VGDAKGYERLRAAHPOAVVILGW--QHNP- 349  
 361 WGPAPRAVILGAVPRGRLVLDLFAESQPVYT-RTASFGQGFITCMJLHNGNGLFGA 419  
 350 -----RAIVDAVYDKRMVLVVDGLCDRFPKVTDEADMGHGPYAFSGIMVFGG-HTLGA 402  
 420 LEAVNGGEAARLF-----PNTMVGTMAREGISQNEVVYSIMAEIGMRKDPVPLAA 473  
 403 -----NTPDMASLTERMTFRPGSTLRGVALPEADNPPAFALFSLARREGDL-DLRA 466  
 474 WTSFAARRYGVSHPDGAAMRLILRSYVNGSGEAGHNRSPLY-RRPSLQMTSTIWN 532  
 457 WFAWARSRYGRDPHAEADWILRRTAYGTTADSMSEGDGLFGARPSLAATKASWS 516  
 533 -----RSDVE-AMRLILTSABSLATSPARVYLLDLTROAVELVSLTYEERKAYLS 585  
 517 PKRLRYRREEFEPALGELLKVRPGLRGSASARRDLVVARQALSNRSRVLIPQIRIAYEA 576  
 586 KELASLIRAGGV-LAYELLPALDEVLASDRFLLGSVLEQARAAVSEADAFVEQSRX 644

## RESULT 4

US-10-156-761-13515  
 ; Sequence 13515, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; PRIOR FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 13515  
 ; LENGTH: 1038  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-13515

Query Match 21.0%, Score 827.5; DB 15; Length 1038;  
 Best Local Similarity 29.9%; Pred. No. 4.1e-66;  
 Matches 222; Conservative 121; Mismatches 359; Indels 41; Gaps 15;

3 AVAANAAGVLLLAGAGGAGD-EAREAAVRLVATLGGPDAADSVSVERALAAKRG 61  
 15 AIGLGTGTCTALPAHATSPGEGPALDTAPASALNRL-PCHAQFRSLT--VGRTRG 70  
 62 LDYSLGGGGAARVRVSGTGVAAAGLHRYLDFCGCHVAMSGSGLRPRPLPAVAGEL 121  
 71 RDRFRV-TGGTRIRIVSGTTPAVLLITGVHWYLVKVCAGHLANNGQLDLPRLPAPAPPL 129  
 122 TEATPNRYRYQNVCTQSYFVWMDMAREREIDMMALNGINILAMSGOEALMORVYLA 181  
 130 ERSTALSHRFALNDNDGTAPYADMSYEHQIDLLAHGNEVAVVINGTEAVHYRLXD 189  
 182 LGLTQAEINEFTGPAFLAMGRMGNLHTWDGRLPPSWHIKOLYLOHRVLDQMSFGMTV 241  
 190 FGYSDETAAMWLPASHPQWMLQNLGSGYGLSPELTAERAGLRRICDRIRALGMAV 249  
 242 LPFAGHVPBAVTVFPQVNVTKMGSGHFNCSYSCSFLAPEDPIPTIGSLFRELIX 300  
 250 LRGYIGHFPGKGVENGGDAHVYPOGIWGF---ERDWDLDFRTASPAAYAKSYRHK 305  
 301 KEFGTDHYGADTENNOMPPSEPSYLAATTAAYEAMTAVDTEAVVLLQGMFOHOPF 360  
 306 DVFGKAHFKMDLHEGTAQDVP--VFGAAGVYKALQAAHPGATVYILGWEANPLP- 361  
 361 WGPAPRAVILGAVPRGRLVLDLFAESQPVYT-RTASFGQGFITCMJLHNGNGLFGA 419  
 362 -----ALDLAIDKKKMLVYDGVSDRYTSYDREKMDGGPIYAFGTIPNFGGRTTIGAR 414  
 420 LEAVNGGEAARLF-----PNTMVGTMAREGISQNEVVYSIMAEIGMRKDPVPLAAVTSFA 479  
 415 AHUMNEKFAWMDKAGSALAGTAYLPEAADRDPAFALFSELAWAGXI-DRAAWFSSYA 473

```

CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEO ID NOS: 1410
SEO ID NO 5107
SOFTWARE: FastSeq for Windows Version 4.0
LENGTH: 4342
TYPE: FRT

```

```

Query Match 3.14; Score 121; DB 9; Length 4342;
Best Local Similarity 21.4%; Pred. No. 1.4;
Matches 178; Conservative 101; Mismatches 318; Indels 236; Gaps 44;

QY      4 VAVAA-----ANGVLLIAGAGG-----AGDEAPAAAVRA-LVRL 40
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1181 VALCAERSPOLLIALLIVRAAGGVFLDEDPSERLAYMLADSGVLLLTQALHIFRL- 1239
QY      41 GPG-----PAADPSVSVERLAAKPGIDTSLGGCGGAARVRGSGTVAALAAGHRFLRD 95
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1240 -PGAEVTPICLDSLKLDNNPQAPGLHLH--GDNLAYVYTTSGSTQPKVG----- 1289
QY      96 FCGCHVMSGSQGLRPRLPAVPE--LTEATPRRYEYGVNCTQSYSPYMDWARERE 153
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1290 --NTHALA-ERLQMQATLTLDGDVLMQAP-----VSFVSVW--E 1348
QY      154 IDWMAALNGINALAMSGO-----EAIQRYVLALGLQAEINEPFTGPAFLAMRG 205
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1329 CFMLVLVGCRLVLAAPBHRDPAVLVELVQFVTLTHFVPELLQLTDEGVAACSLR 1388
QY      206 NHTWDPRLPSMHIKQLYLQHRFLDQMSFGMTPLPAFAGHVEAVTRPFPQVNTYKM 265
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1389 RLSEGGALPPE-----LRNVLQR-----LPAALAHRYGTER--AINTY 1428
QY      266 GSWCHENCSSCSFLAPD---PIPPIIGLFLRELKEFGTDHYGADTENEM--- 317
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1429 --HMQC-----RAEDGERSPIGRPLGNVVCVILDAEF---NLLPGVAGELTIGG 1473
QY      318 -----QPPSEPEYLAATNAVYEAHTAVDEAVMLQGL-----FQHQPOKMG- 362
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1474 LGLARGYLGRFALSAERFVADPFSAAGERLYRTGDRARMADGLEYLRGLDQVYKRGF 1533
QY      363 ---PAQIRAVYGAVP---RGRLLVLDFFAESQPY--YTRTASFOGQEPFWCMLENFGNH 414
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1534 RIEBEELQARLLAQPGVAAVVRBEGAASQLVGYTTGAVGAABA-----EQNQ 1584
QY      415 GLFALBAVNGGPPAARLPFNSTVNTGTMAPEGISONEVYVSLMAELGMK---DVPD 470
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1585 RLRAALDA---ELPEYMWPTQIMRLACMPLGPS--GKLTJRALPEPWOQREHYEPTE 1638
QY      471 L-----AAWTSFAARYG-----VSHDDAQAARLLIRSYNCS- 505
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1639 LQRIALIMSVGLGLPRVGLDDPFEELGSHSLATRLVSRTRQCDVELPLRALFEASEL 1698
QY      506 -----GEACHGNRSLPVRRSPLOMTSTIWNRSVDFEAWRLLLTSAPLSTSPAF 556
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1699 EAFCEQYRAAQAGRTDSHGAIIRIDREQRPVLSYQRMFLWLQ-----EPD--SPAY 1751
QY      557 RY-----DLIDLTR--QAVQELVSLYEABARAYLSKELASLLRA---GGV-LAYELL 603
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 1752 NVGLIARLSPGLDVAFREAFALQALVQ-RHETLRTTPSPVDGVQVHGGGLHMDWQDF 1810  
 Qy 604 PALD-----EVLASRSRFLGSMLEQA---PAAVSEKADPFYQNSRYQULTLMGPE 652  
 Db 1811 SALDRSRQOHLQTLA-DSEAHKRPDLSEGLPLVCVAKAEREHYLVTLHIV---TE 1866  
 Qy 653 GNLDYANKQGLAVANYTTPRWRLFLALVDVSAQGIPEQOHPQKXVFOLE 705  
 Db 1867 GMANDIFARELGLAYEAFLDNR-----ESPLFELPLVQ-YLDVSVQRE 1908

RESULT 7  
 US-10-156-761-9695  
 ; Sequence 9695, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156, 761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 9695  
 ; LENGTH: 469  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-9695

Query Match 3.0%; Score 118.5; DB 15; Length 469;  
 Best Local Similarity 22.1%; Pred. No. 0.078;  
 Matches 126; Conservative 49; Mismatches 167; Indels 229; Gaps 32;  
 Qy 208 HTWDDPLPSPWHIKOLYLOHRYLDM-----RSFMTPTVLPAPA 246  
 Db 14 HEMKCALPP-----LIRPTVDELARAAALVARRPGDKRLRVGTSGKPTLMILSV 66  
 Qy 247 GH-VPEAVTRVFPVQVNTTKMGWGHFNCSYSCFLAPEDPIPIIGSLFLRELKEF-G 304  
 Db 67 GHGTRHALVAGPHANE-----PVGATALRLAERYLAG 100  
 Qy 305 TDHITYGATFEMQPPSEPSYLAATTAVYEAMTAVDTEAVMLLQGL-----FQ 355  
 Db 101 PGPAGAD-----ATNNLLICDDPGARNEGSLSPYTLGHYFR 140  
 Qy 356 H-----QPFMGPAQIRAVLGAVPRGRLVLDFAESQPVYTRTASFGQGFPIWCM 407  
 Db 141 HFFRRGFLQEPWLPDGAERA---ALPSTRAL-LDLODEL-----REFPQCSL 184  
 Qy 408 H-ANGG-----NHSLFALFAVNGGPPAAR-----LFPNSIMVGTGMAPEGISQBEV 454  
 Db 185 HGVVDGGGFFVELTODLPGLAQIVAG--TPAARFGIPREIGPYDTLWPPPLG----AV 235  
 Qy 455 YSLMELGMRKDPV--DLAAWTS-----FAARRGV------S 486  
 Db 236 YRI-----PPRRGDILAAITFAVESTWTFHPQKYGTVTAVVEAPMWGAAYEDGS 286  
 Qy 487 HP-DGAAMRLILRRVNVCSG--EACRGHNSPLVRRPSLQMTSITWNRSDVEARLL 543  
 Db 287 PFGDADVALRAVSRLRDGTGLIEDILARIRPRLAATPD---TASLAPVDD---Y 336  
 Qy 544 LTSAPSLATSPAFRDLDLTFQAVQEL-----VSLYEARSAYLSKELASLLRA 594  
 Db 337 LLVGVGLADS--WDDTFTDGTGRPLPLPNTARLTLGSGRRVALKRTGLRLQALAT---A 391

Qy 595 GGVALYELLPALDEVILASDSRFLGSMLEQAARAAVSEADPFYQNSRYQULTLMGPEGN 654  
 Db 392 AGAASDVLPEDNR-----LIDEMCADYI-----DGYAR-----WVPVGH 427  
 Qy 655 ILDYANK-----QLAGLVANYTT---PRW 675  
 Db 428 OVEYQTRVVLAAFAELAGRYAPVRSHSGEPRW 458

RESULT 8  
 US-10-307-019-7  
 ; Sequence 7, Application US/10307019  
 ; Publication No. US20030108533A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: zeng, Wenlin  
 ; APPLICANT: Stanton, Lawrence  
 ; APPLICANT: SCIOS, INC.  
 ; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION  
 ; FILE REFERENCE: SCIOS.021DVL  
 ; CURRENT APPLICATION NUMBER: US/10/307,019  
 ; CURRENT FILING DATE: 2002-11-26  
 ; PRIOR APPLICATION NUMBER: 09/548,473  
 ; PRIOR FILING DATE: 2000-04-13  
 ; PRIOR APPLICATION NUMBER: 60/129,552  
 ; PRIOR FILING DATE: 1999-04-16  
 ; NUMBER OF SEQ ID NOS: 9  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 7  
 ; LENGTH: 871  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-307-019-7

Query Match 3.0%; Score 118.5; DB 15; Length 871;  
 Best Local Similarity 18.6%; Pred. No. 0.2;  
 Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;  
 Qy 19 GGNAGDEARBAAVRALVRL-LGQGPAD---FSVSRALAAAPGLD-----TY 65  
 Db 21 GSAQSQSSQSVSLRGSSQVGTSPSLDAGGWTQEADLSDSTPTLQRPQEQATMKXF 80  
 Qy 66 SLGG-CGAARVVRGSGTVAAAGLHRYLRDCG-----CHVAWSQSQR----- 109  
 Db 81 SLGGRGTVAGVAGYGFATGCDAG-----GMLGGPMWARIAMVSSQSEEEQEARAE 134  
 Qy 110 -----LP---RPLAVPGELTATPRRYQNVCTQSYSPWMDMAWE 151  
 Db 135 SQSEEQEARASPLPQVSARFVPEVGRAPTRSSP-----EPTWE 175  
 Qy 152 -----RET--DWMALNGINLAL-----AMSGQBAIMQRYIAL----- 182  
 Db 176 DIGQVSLVQIRLDSGDAEADTISLDSISVDAYLNLSDLYIKTLFPFMIFRVPKSA 235  
 Qy 183 -----GLTQAEINEF-----FTGP-----ATLWGRMGNLHTWGP 213  
 Db 236 QPEPSPMAEEELAEFPETWMPGELGPHAGLETTESEEDVDALAEAAVGRKRWSSP 295  
 Qy 214 LPPSWHILKOLYLOHRYLDMRSFGMTPVLPAPAGHV-----PEAVTRVFPQVNTTKMG 266  
 Db 296 SRSLLFFPG--RHLPDPAELGLERKVASVEHISRLIKRPFGLKEGEP---RKRP 349  
 Qy 267 SWGHFNCSYSCFLAPEDPIPIIGSLFLRELKEFGTDHLYGADTFEMQPPSEPSY 326  
 Db 350 GLASFRLGLKSWDRAP-----FLREL-----SDETVV 378  
 Qy 327 LAAATTAVEAMTAVDTEAVMLLQGLPFGHQGFQFNGPAIRAVLGAVPG--RLVLDLF 384  
 Db 379 LGQSVTLAQVSAQPAQATWCKDAPLE-----SSKRLISATLTKNPFLLILVVV 430  
 Qy 385 AESQPVYTRTASFOGPFITWCMILNFGGNHGLFGALLEAVNGGPEARLFPNSITWVGTA 444  
 Db 431 AEDLVGYT-----GVSNALGVTVTGVLK---KAR--PSSS----- 463

QY 445 PEGISQNEVVYSLMAELGWRKDPVPD-----LAAM--VTSFAARRGVSHPDGAA 493  
 Db 464 -----PCPDIGEVYADGVLLVWKPVESYGVTYIVQCSLEGS 501  
 QY 494 WRLLRSVYNC-----SGEAC 509  
 Db 502 WTLASDIFDCCYLTSKLSRGTYTFRACVSKAGMPYSSPSBQVLLGSPSHLASEES 561  
 QY 510 RGNHRSPLYRRPSIQMTSTMYNR-SVFEAMRLLLTSAFSLATSAPFYDLDLTR--Q 566  
 Db 562 QGRSAQPLPSTKTFAFOTQIORGRFSVVRQCEKASGRALAAKTIYPHPKDTAVLREYE 621  
 QY 567 AVOELVSLYYEARSAYLS-KELASLRAGVLAVELLPALDEVLASDSRFLIGSWLEQA 625  
 Db 622 ALKGLRHPHLAQHLAAVLSPRHLVLIJEL--CSGPELLPCLAE-RASYSESEVQDYLMQM 678  
 QY 626 RAAA-----VSEADRYEQN---SRQYLTIMGPEGNILIDYANKQLAGLVANYTPRML 677  
 Db 679 LSAQYVLNQHILHLDIRSENMTITEYNLL-----KVVDLGNAGSLSGEKVLPSPDKFO 732  
 QY 678 FLEALVDSVAG--IPFOQHOPDKNVFOL-EQAFVLSKQRYSPQPRG 721  
 Db 733 YLEMAPLELLEGQAVP-----QTDIMAGVTAFTMLSAEYVSSSG 774

RESULT 9  
 US-10-307-019-1  
 ; Sequence 1, Application US/10307019  
 ; Publication No. US20030108533A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zeng, Wenlin  
 ; APPLICANT: Stanton, Lawrence  
 ; APPLICANT: SCIOS, INC.  
 ; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION  
 ; FILE REFERENCE: SCIOS.021DVL  
 ; CURRENT APPLICATION NUMBER: US/10/307, 019  
 ; CURRENT FILING DATE: 2002-11-26  
 ; PRIOR APPLICATION NUMBER: 09/548,473  
 ; PRIOR FILING DATE: 2000-04-13  
 ; PRIOR APPLICATION NUMBER: 60/129,552  
 ; PRIOR FILING DATE: 1999-04-16  
 ; NUMBER OF SEQ ID NOS: 9  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 1351  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-307-019-1

Query Match 3.0%; Score 118.5; DB 15; Length 1351;  
 Best Local Similarity 18.6%; Pred. No. 0.39;  
 Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;  
 QY 19 GGAAGDEAREAAVAVLVARL-LGGPAD--FSVSVERALAKFGD-----TY 65  
 Db 501 GSAAGASSQVSSIRVSSQVTEPGPSLDAGWTQAEADLSSTPTLQRPQEQATMKRF 560  
 QY 66 SLGG-GGAARVTVGSGVAAAAGLHRYLADFCG-----CHVAMSGQLR-----109  
 Db 561 SLGGRGYAGVAGYGTAFGGDAG-----GMLGGQWMAKRLMAYVSSQSEEEQEARAE 614  
 QY 110 -----LP---RPLPAVGELEATPNNRYRYQVNCQSYSFVWMDARWE 151  
 Db 615 SQSEEQEARAESPLPOVSARPYVEVGAPTRSSP-----EPTPW 655  
 QY 152 -----RET---DMMLNGINTAL-----AMSGEATWQVYVLL-----182  
 Db 656 DIGVSLVQIRDLSDGDAADTISLDISEVDPAVLNLDYIKYLPFEPMIRKYPKSA 715  
 QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGRMGLMTWDGP 213  
 Db 716 QPSPSMABEELAEFPEPTWPGELGPHAGLEITERSESDVALLAEAVAGKRWSSP 775

QY 214 LPSPMHLKQLYLQHRVIDQMSFGMTPEVLPFAFAGV-----PEAVTRVFPQVNVTKMG 266  
 Db 776 SRSLEHPPG---RLPLDPERELGLRRVYASVHERILIKGREGEKEGCP--RKP 829  
 QY 267 SMGHNCSYSCSFLAABEDPLFPLIGSLFRELKEGTGHIIYADTFNEMQPPSSPSY 326  
 Db 830 GLASFRLSGLXSWDRAP-----TFLEL-----SDETVV 858  
 QY 327 LAAATTAVEAMTAVTEAVVLLQGMLEFOHQPFCWGPQIRAVAGVPRG--RLVLDLF 384  
 Db 859 LGQSTTACQVSAAPPAQATMSKQAPLE-----SSRVLISATLKNQVLLTVV 910  
 QY 385 AESQPVYTRTASFQGPPTWCMNLNFGNHLFGALVAVNGGPPAARLFPNSTWVGWMA 444  
 Db 911 AEDIGVY-----CSVSNALGTVTTGVLR-----KAER--PSSS-----943  
 QY 445 PEGISQNEVVYSLMAELGWRKDPVPD-----LAAM--VTSFAARRGVSHPDGAA 493  
 Db 944 -----PCPDIGEVYADGVLLVWKPVESYGVTYIVQCSLEGS 981  
 QY 494 WRLLRSVYNC-----SGEAC 509  
 Db 982 WTLASDIFDCCYLTSKLSRGTYTFRACVSKAGMPYSSPSBQVLLGSPSHLASEES 1041  
 QY 510 RGNHRSPLYRRPSIQMTSTMYNR-SVFEAMRLLLTSAFSLATSAPFYDLDLTR--Q 566  
 Db 1042 QGRSAQPLPSTKTFAFOTQIORGRFSVVRQCEKASGRALAAKTIYPHPKDTAVLREYE 1101  
 QY 567 AVOELVSLYYEARSAYLS-KELASLRAGVLAVELLPALDEVLASDSRFLIGSWLEQA 625  
 Db 1102 ALKGLRHPHLAQHLAAVLSPRHLVLIJEL--CSGPELLPCLAE-RASYSESEVQDYLMQM 1158  
 QY 626 RAAA-----VSEADRYEQN---SRQYLTIMGPEGNILIDYANKQLAGLVANYTPRML 677  
 Db 1159 LSAQYVLNQHILHLDIRSENMTITEYNLL-----KVVDLGNAGSLSGEKVLPSPDKFO 1212  
 QY 678 FLEALVDSVAG--IPFOQHOPDKNVFOL-EQAFVLSKQRYSPQPRG 721  
 Db 1213 YLEMAPLELLEGQAVP-----QTDIMAGVTAFTMLSAEYVSSSG 1254

RESULT 10  
 US-10-307-019-4  
 ; Sequence 4, Application US/10307019  
 ; Publication No. US20030108533A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zeng, Wenlin  
 ; APPLICANT: Stanton, Lawrence  
 ; APPLICANT: SCIOS, INC.  
 ; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION  
 ; FILE REFERENCE: SCIOS.021DVL  
 ; CURRENT APPLICATION NUMBER: US/10/307, 019  
 ; CURRENT FILING DATE: 2002-11-26  
 ; PRIOR APPLICATION NUMBER: 09/548,473  
 ; PRIOR FILING DATE: 2000-04-13  
 ; PRIOR APPLICATION NUMBER: 60/129,552  
 ; PRIOR FILING DATE: 1999-04-16  
 ; NUMBER OF SEQ ID NOS: 9  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1610  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-307-019-4

Query Match 3.0%; Score 118.5; DB 15; Length 1610;  
 Best Local Similarity 18.6%; Pred. No. 0.51;  
 Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;  
 QY 19 GGAAGDEAREAAVAVLVARL-LGGPAD--FSVSVERALAKFGD-----TY 65  
 Db 760 GSAAGASSQVSSIRVSSQVTEPGPSLDAGWTOAEADLSSTPTLQRPQEQATMKRF 819



```

QY 66 SLGG-GGAARVRVNGSTGVAAGLHRYLADFCG-----CHVAMSGSOLR-----109
DB 820 SLGGGAGVAGVAGYGTAFPGDAG-----GMLGGGMMARIANAASQSEEEGEARAE 873
QY 110 -----LP-----RFLPAVPGELTEATPNRYRYQNVCTOSYSFVMMDMARME 151
DB 874 SQSEGEQEARABESLPQVSARFPEVGRAPTRSSP-----EPTPWE 914
QY 152 -----REI--DMALINGINAL-----AMSGEAIHQRYLAL-----182
DB 915 DIGQVSLVQIRDLGDAADTISLDISEVDPAVNLSDLDYDKYLPFEPMIFRKYKSA 974
QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGRMGLHTWDGP 213
DB 975 QPEPPSPMAEBELAFPEPTWMPGELGPHAGLEITESEVDALLAEAVGRKKMSSP 1034
QY 214 LPPSMHKKQLYLQHRVLDQMSFGMTVLPAPAGHV-----PEAVTRVFPQVNTVMG 266
DB 1035 SRSLEHFPFG--RHLPDEPAELGLRERKASVEHISILKGRPEGLEKGGP--RKKP 1088
QY 267 SMGHFNCSYSCSFLAPEDPIFPIIGSLFLRELKERGTDHIYADTFNEMQPPSEPSY 326
DB 1089 GLASFRLSGLKSWDAP-----TFIREL-----SDETIV 1117
QY 327 LAAATTAVYEAMTAVDTEAVWLQGMWLFQHQPFQMPGAPQIRAVLGAVERG--RLVLDLF 384
DB 1118 LGQSVTLACQVSAQPAQATWSKDGAPE-----SSSRVLISATLKNFQLTLLVVV 1169
QY 385 AESQPVYTRTASFGQGFPIWCMHNFGNHLFGALAVNGGPEARLFPNSTWVTGMA 444
DB 1170 AEDLCVIT-----CSVSNAIGVTYTTGVLR-----KAER--PSSS-----1202
QY 445 PEGISONEVYVSLMAELGMRKDPVD-----LAAM--VTSFAARRYGVSHPDAGAA 493
DB 1203 -----PCPDIGEVYADVGLVWKPVEISYGPVTYIVQCSLEGGS 1240
QY 494 WRLLRSVYNC-----SSEAC 509
DB 1241 WTTLASDIFDCCYLTSKISRGTYTFTACVSKAGMGPSSPSBOVLLGAPSHLASSEES 1300
QY 510 RGHNSPLVRRSPLOMNTSIWYNR--SDVEAMRLLLTSPSLATSPAFRYDLDTLR--Q 566
DB 1301 QGRSAQPLPSTKTPAFQIQGRFVSVRQCEKASGRLAAKIIIPHPKDXAVLREYE 1360
QY 567 AVQELVSLYEBARSAIYS--KELASLLRAGGVLAYELLPALDEVLASDSRFLGSMLEQA 625
DB 1361 ALKGLRHPHLAQLHAAYLSPRHVLITEL--CSGPELLPCLAE--RASYSSEVVDYLMQM 1417
QY 626 RAAA-----VSEAEADFEYQ--SRYQLTLMGPEGNIIDYANKQLAGLVANYTTPRML 677
DB 1418 LSAQVYLHNQHLHLDLSENNIITEYNLL-----KVVDLGNAGSLSOEKVLPBDKXKD 1471
QY 678 FLEALVDSVAG--IPFQGHQFDKXVQOL--EQAVLSKORPSQPRG 721
DB 1472 YLETMAPELLBQGAVP-----QTDIMAIGVTAIFIMLSAEYVPSSEG 1513

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-858-664A-2
Query Match      3.0%; Score 118.5; DB 9; Length 1665;
Best Local Similarity 18.6%; Pred. No. 0.54;
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

QY 19 GGAADDEAREAAVALVARI--LGPAPAD--FSVSVERALAAKPGLD-----TY 65
DB 815 GSASQASSSQVSLVVGSSQVTEBPGLDAGMTQAEEDLSSTPTLQRPQEQVTMRKF 874
QY 66 SLGG-GGAARVRVNGSTGVAAGLHRYLADFCG-----CHVAMSGSOLR-----109
DB 875 SLGGGAGVAGVAGYGTAFPGDAG-----GMLGGGMMARIANAASQSEEEGEARAE 928
QY 110 -----LP-----RFLPAVPGELTEATPNRYRYQNVCTOSYSFVMMDMARME 151
DB 929 SQSEGEQEARABESLPQVSARFPEVGRAPTRSSP-----EPTPWE 969
QY 152 -----REI--DMALINGINAL-----AMSGEAIHQRYLAL-----182
DB 970 DIGQVSLVQIRDLGDAADTISLDISEVDPAVNLSDLDYDKYLPFEPMIFRKYKSA 1029
QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGRMGLHTWDGP 213
DB 1030 QPEPPSPMAEBELAFPEPTWMPGELGPHAGLEITESEVDALLAEAVGRKKMSSP 1089
QY 214 LPPSMHKKQLYLQHRVLDQMSFGMTVLPAPAGHV-----PEAVTRVFPQVNTVMG 266
DB 1090 SRSLEHFPFG--RHLPDEPAELGLRERKASVEHISILKGRPEGLEKGGP--RKKP 1143
QY 267 SMGHFNCSYSCSFLAPEDPIFPIIGSLFLRELKERGTDHIYADTFNEMQPPSEPSY 326
DB 1144 GLASFRLSGLKSWDAP-----TFIREL-----SDETIV 1172
QY 327 LAAATTAVYEAMTAVDTEAVWLQGMWLFQHQPFQMPGAPQIRAVLGAVERG--RLVLDLF 384
DB 1173 LGQSVTLACQVSAQPAQATWSKDGAPE-----SSSRVLISATLKNFQLTLLVVV 1224
QY 385 AESQPVYTRTASFGQGFPIWCMHNFGNHLFGALAVNGGPEARLFPNSTWVTGMA 444
DB 1225 AEDLCVIT-----CSVSNAIGVTYTTGVLR-----KAER--PSSS-----1257
QY 445 PEGISONEVYVSLMAELGMRKDPVD-----LAAM--VTSFAARRYGVSHPDAGAA 493
DB 1258 -----PCPDIGEVYADVGLVWKPVEISYGPVTYIVQCSLEGGS 1295
QY 494 WRLLRSVYNC-----SSEAC 509
DB 1296 WTTLASDIFDCCYLTSKISRGTYTFTACVSKAGMGPSSPSBOVLLGAPSHLASSEES 1355
QY 510 RGHNSPLVRRSPLOMNTSIWYNR--SDVEAMRLLLTSPSLATSPAFRYDLDTLR--Q 566
DB 1356 QGRSAQPLPSTKTPAFQIQGRFVSVRQCEKASGRLAAKIIIPHPKDXAVLREYE 1415
QY 567 AVQELVSLYEBARSAIYS--KELASLLRAGGVLAYELLPALDEVLASDSRFLGSMLEQA 625
DB 1416 ALKGLRHPHLAQLHAAYLSPRHVLITEL--CSGPELLPCLAE--RASYSSEVVDYLMQM 1472
QY 626 RAAA-----VSEAEADFEYQ--SRYQLTLMGPEGNIIDYANKQLAGLVANYTTPRML 677
DB 1473 LSAQVYLHNQHLHLDLSENNIITEYNLL-----KVVDLGNAGSLSOEKVLPBDKXKD 1526
QY 678 FLEALVDSVAG--IPFQGHQFDKXVQOL--EQAVLSKORPSQPRG 721
DB 1527 YLETMAPELLBQGAVP-----QTDIMAIGVTAIFIMLSAEYVPSSEG 1568

```

RESULT 12

US-10-307-019-6  
; Sequence 6, Application US/10307019  
; Publication No. US2003010853A1  
; GENERAL INFORMATION:  
; APPLICANT: zeng, Wenlin  
; APPLICANT: Stanton, Lawrence  
; APPLICANT: SCIOS, INC.  
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION  
; FILE REFERENCE: SCIOS.021DVI  
; CURRENT APPLICATION NUMBER: US/10/307,019  
; PRIOR FILING DATE: 2002-11-26  
; PRIOR APPLICATION NUMBER: 09/548,473  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/129,552  
; PRIOR FILING DATE: 1999-04-16  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 2596  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-307-019-6

Query Match 3.0%; Score 118.5; DB 15; Length 2596;  
Best Local Similarity 18.6%; Pred. No. 1.1;  
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

19 GGAAGDEAREAAVRLVRL-LGPGPAD---FSVVERALAAKPGD-----TY 65  
1746 GSAQASSSQVSSLRVSSQVGTPEPGLDAGMTQEAEDLSDSTTLRPOEQATWRKF 1805  
66 SLGG-GGAARVRVGTSTGVAAAAGLHRYLRDFCG---CHVAMSSGQLR----- 109  
1806 SLGGRGVAGVAGYGTFAFGDAG---GMLGGPMARIAMAVSQSEEEQEARAE 1859  
110 -----LP-----RPLPAVPGELTEATPRHYRYQNVCTQSFSFWMMDARWE 151  
1860 SQSEEQEARAESPLPVQSARFVPEVGRAPTRSSP-----EPTPWE 1900  
152 -----REI---DMMALNGINLAL---AMSGOEAIWQRYVAL----- 182  
1901 DIGQVSLVQIRLDGDAEADTISLDISEVDPAVYMLSDLYDIKYLPEFMIFRKVKESA 1960  
183 -----GLTQAEINER-----FTGP-----AFLANGRMGNLHTWDGP 213  
1961 QPEPSPMAEBELAEPEPTWMPGELGPRAGLEITESEBDVALLAEAVGKRRKSSP 2020  
214 LPPSWHLKQLYLQHRVLDQKRSFGMTPVLPAPAGHV-----PEAVTRVFPQVNVTKMG 266  
2021 SRSLFHPFG---RHLPLDEPAELGLRERYKASVEHISRLIKGRPEGLEKGGP---RKKP 2074  
267 SMGHFNCSYSCSFLAPEDPIFPIIGSLFLRELIKRGTGDIHYGADTFNEMQPSSEPSY 326  
2075 GLASFRLSGKSWDRAP-----TFRLREL-----SDETTV 2103  
327 LAATTAVYEAAMTAVDEAVVLLQGMFLQHPQFQWGPAPQIRAVLAGVPRG---RLVLDLF 384  
2104 LGQSVTLACQVSAQPAQAATWSKDGALE-----SSSRVLISATLKNFOLLITLVV 2155  
385 AESQPVYTRTASFQGGPFIVCMLEHFGNHLFGALEAVNGGEAARLPNSTMGVTGMA 444  
2156 AEDIGVYT---CSVSNALGTVTTGVLR-----KABR---PSSS----- 2188  
445 PEGSQSQUEVVYSIMABELGMRKDPDP-----LAAM---VTSFAARRGVHPDAGAA 493  
2189 -----PCPDIGEVYADGVLLVWKPVSYGPPVYIYQCSLEGG 2226  
494 WRLLRSVYNC-----SGEAC 509  
2227 WTLIASDIFDCCVLTSLKSRGTYTFRACVSKAGMGPYSPSEQVILGAPSHLAEBS 2286  
510 RGNHRSILVRRPSQWNTSITWNR-SDFEAWRLLLTSAPSLATSPAFRDLIDLTR--Q 566

DB 2287 QGRSAQPLPSTKTFAPQTOQRGRFSVWQWEKASGRALAAKLIPIHPKDKTAVLREYE 2346  
QY 567 AVQELVSLYEARSAYLS-KEIASLIRAGCVLAIYELLPALDEVLASDSRFLIGSLQEA 625  
DB 2347 ALKGLRHPHQAQHAALVSPRHVLTLEL---CSGPILLPOLAE-RASYSESEVKDLMQW 2403  
QY 626 RAAA-----VSEADAEYQON---SRVQLTMWPEGNILDYANKQLAGLVANYTPRRL 677  
DB 2404 LSAQVYHNGHILHLDLRSENMILTEYNLL-----KVVDLGAQSLQSEKVLPSDKFQ 2457  
QY 678 FLEALYDSVAQ---IPQOHQPDKNVFOU-EOAFVLSKORYSQPRG 721  
DB 2458 YLETMABELLEGQAVP-----QTDIMAGVAFATMLSAEYVSSEG 2499

RESULT 13  
US-10-077-130-2  
; Sequence 2, Application US/10077130  
; Publication No. US20020168742A1  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan L.  
; APPLICANT: Kapeller-Libermann, Rosana  
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family  
; FILE REFERENCE: MPI2001-04/PIKCP1(W)  
; CURRENT APPLICATION NUMBER: US/10/077,130  
; PRIOR FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 60/269201  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 2630  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-077-130-2

Query Match 3.0%; Score 118.5; DB 14; Length 2630;  
Best Local Similarity 18.6%; Pred. No. 1.1;  
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

19 GGAAGDEAREAAVRLVRL-LGPGPAD---FSVVERALAAKPGD-----TY 65  
1780 GSAQASSSQVSSLRVSSQVGTPEPGLDAGMTQEAEDLSDSTTLRPOEQATWRKF 1839  
66 SLGG-GGAARVRVGTSTGVAAAAGLHRYLRDFCG---CHVAMSSGQLR----- 109  
1840 SLGGRGVAGVAGYGTFAFGDAG---GMLGGPMARIAMAVSQSEEEQEARAE 1893  
110 -----LP-----RPLPAVPGELTEATPRHYRYQNVCTQSFSFWMMDARWE 151  
1894 SQSEEQEARAESPLPVQSARFVPEVGRAPTRSSP-----EPTPWE 1934  
152 -----REI---DMMALNGINLAL---AMSGOEAIWQRYVAL----- 182  
1935 DIGQVSLVQIRLDGDAEADTISLDISEVDPAVYMLSDLYDIKYLPEFMIFRKVKESA 1994  
183 -----GLTQAEINER-----FTGP-----AFLANGRMGNLHTWDGP 213  
1995 QPEPSPMAEBELAEPEPTWMPGELGPRAGLEITESEBDVALLAEAVGKRRKSSP 2054  
214 LPPSWHLKQLYLQHRVLDQKRSFGMTPVLPAPAGHV-----PEAVTRVFPQVNVTKMG 266  
2055 SRSLFHPFG---RHLPLDEPAELGLRERYKASVEHISRLIKGRPEGLEKGGP---RKKP 2108  
267 SMGHFNCSYSCSFLAPEDPIFPIIGSLFLRELIKRGTGDIHYGADTFNEMQPSSEPSY 326  
2109 GLASFRLSGKSWDRAP-----TFRLREL-----SDETTV 2137  
327 LAATTAVYEAAMTAVDEAVVLLQGMFLQHPQFQWGPAPQIRAVLAGVPRG---RLVLDLF 384  
2138 LGQSVTLACQVSAQPAQAATWSKDGALE-----SSSRVLISATLKNFOLLITLVV 2189

```

QY 385 AESQVYTRTASFOGQPFIMCMLHNFQGNHGLFGALEAVNGPPEARLFPNSTVWGTGMA 444
| | | | |
Db 2190 AEDGVTT-----CSVSNALGTVTTTVLR-----KAER--PSSS----- 2222
QY 445 PEGISQNEVYVSYSLMAELGWRDVPVD-----LAAM--VTSPFAARRYGVSHPDAGAA 493
| | | | |
Db 2223 -----PCPDIGEVYADVLLVWKPEVSEYGVITYVQCSLEGGS 2260
QY 494 WRLLRSYVNC-----SEAC 509
| | | | |
Db 2261 WTTLASDIFDCCYLTSTKLSRGCTYTFRTACVSKAMGPSSPSEQVLLGSPSHLASEES 2320
QY 510 RGNHNSPLVRRPSLOMTSITWYNR--SDVFEAMRLLTSPSLATSFAFRYDLDLTR--Q 566
| | | | |
Db 2321 QGRSAQPLPSTKTFAPQIQGRGFSVVRQCEKASGALAAKIIIPHPKDXAVLRBYE 2380
QY 567 AVQELVSLYEBARSAIYS--KELASLRAAGVLAELLPALDEVLASRFLGSMLEQA 625
| | | | |
Db 2381 ALKGLRHPHLAQLHAAYLSPHLLVLIHEL--CSGPELLPCLAE--RASYSSEVKDYLMQM 2437
QY 626 RAAA-----VSEAEADFYEQN--SRVQTLTGPEGNILDYANKQLAGLVANYTTPRWRL 677
| | | | |
Db 2438 LSAQVYLANQHILHLDLASEMITEYNLL-----KVVDLGNQSLSQEKVLPBDEKXD 2491
QY 678 FLEALVDSVAQG--IPFOHQFDKNVFOU--EQAFVLSKORYPSQPRG 721
| | | | |
Db 2492 YLETWAPPELLGQGAVP-----QTDIWAIGVAFIMLSAEYVSSEG 2533

RESULT 14
US-10-077-130-5
; Sequence 5, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: WPI2001-047P1CPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 7968
; ORGANISM: Homo sapiens
US-10-077-130-5

Query Match 3.0%; Score 118.5; DB 14; Length 7968;
Best Local Similarity 18.6%; Pred. No. 5.8;
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

QY 19 GGAAGDEBARAAAVRALVARL-LGPGPAD--FSVSEVERALAAKGLD-----TY 65
| | | | |
Db 7118 GSAASQASSQVSSLRVSSQVGTPEGPSLDAEGMTQAEHLSTPTLQPGQATMKKF 7177
QY 66 SLUGS-GGAARVVRGSGTGAAGAAQLHRYLRDPG--CHVAMSGQLR----- 109
| | | | |
Db 7178 SLGGRGVAGVAGYGFAPFGDAG-----GMLGQGMARIMAAVSSQSEEBQBARAE 7231
QY 110 -----LP--RPLPAVPGELTEATPNRYRYQVNCQVSFVWMDARWE 151
| | | | |
Db 7232 SQSEEQEARAESELPVYSARVPEVGRATRRSP-----EFTPE 7272
QY 152 -----RET--DWMALNGINAL--AMSQEAIMQSVIAL----- 182
| | | | |
Db 7273 DIGGVSLVQIRLDSGAADATISLDSIVDPAYLINSIDYIKYLPFEEMIRKVPKSA 7332
QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGMLHTWDDG 213
| | | | |

```

```

Db 7333 QDEPPSPMAEBELAPPEPTMWPGEGLPHAGLETTESESDVALLAEAVGRKKWSP 7392
QY 214 LPPSWHILQVLYQHRVLDQMSFGKTVLPAPAGV-----PEAVTRVFPQVNTVQMG 266
| | | | |
Db 7393 SRSLFHPFG--RHLPDEPPELGLRERVKASVEHISILKRPGLKEGEP--RKXP 7446
QY 267 SMGHFNCSYSGFLAPDEDPFPIIGSLFLRELKEFGTDHYGADTFNEMQPSSEPSY 326
| | | | |
Db 7447 GLASFRLSGLSKNDAP-----TFREL-----SDEIVV 7475
QY 327 LAAATVAYEANTAVDTEAVVLLQGMFLFOHQPOFNGPAQIRAVIGAVPRG--RLIVDLF 384
| | | | |
Db 7476 LQQSVTLAQVYSAQPAQATWSKDGAPLE-----SSSRVLSATLKNQQLITLVV 7527
QY 385 AESQVYTRTASFOGQPFIMCMLHNFQGNHGLFGALEAVNGPPEARLFPNSTVWGTGMA 444
| | | | |
Db 7528 AEDGVTT-----CSVSNALGTVTTTVLR-----KAER--PSSS----- 7560
QY 445 PEGISQNEVYVSYSLMAELGWRDVPVD-----LAAM--VTSPFAARRYGVSHPDAGAA 493
| | | | |
Db 7561 -----PCPDIGEVYADVLLVWKPEVSEYGVITYVQCSLEGGS 7598
QY 494 WRLLRSYVNC-----SEAC 509
| | | | |
Db 7599 WTTLASDIFDCCYLTSTKLSRGCTYTFRTACVSKAMGPSSPSEQVLLGSPSHLASEES 7658
QY 510 RGNHNSPLVRRPSLOMTSITWYNR--SDVFEAMRLLTSPSLATSFAFRYDLDLTR--Q 566
| | | | |
Db 7659 QGRSAQPLPSTKTFAPQIQGRGFSVVRQCEKASGALAAKIIIPHPKDXAVLRBYE 7718
QY 567 AVQELVSLYEBARSAIYS--KELASLRAAGVLAELLPALDEVLASRFLGSMLEQA 625
| | | | |
Db 7719 ALKGLRHPHLAQLHAAYLSPHLLVLIHEL--CSGPELLPCLAE--RASYSSEVKDYLMQM 7775
QY 626 RAAA-----VSEAEADFYEQN--SRVQTLTGPEGNILDYANKQLAGLVANYTTPRWRL 677
| | | | |
Db 7776 LSAQVYLANQHILHLDLASEMITEYNLL-----KVVDLGNQSLSQEKVLPBDEKXD 7829
QY 678 FLEALVDSVAQG--IPFOHQFDKNVFOU--EQAFVLSKORYPSQPRG 721
| | | | |
Db 7830 YLETWAPPELLGQGAVP-----QTDIWAIGVAFIMLSAEYVSSEG 7871

RESULT 15
US-09-820-788-4
; Sequence 4, Application US/09820788
; Publication No. US20030106182A1
; GENERAL INFORMATION:
; APPLICANT: SHAO, Wei et al
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
; FILE REFERENCE: CL001207
; CURRENT APPLICATION NUMBER: US/09/820,788
; PRIOR FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Human
US-09-820-788-4

Query Match 3.0%; Score 117.5; DB 12; Length 497;
Best Local Similarity 21.7%; Pred. No. 0.11; Indels 153; Gaps 22;
Matches 97; Conservative 53; Mismatches 144;

QY 345 AVMLL-----QGMFLQHQPOFNGPAQIRAVIGAVRGLVLDLFAESQVYTRTAS 396
| | | | |
Db 14 AIFLLVDLMRRQRMARYPF--GFLPLPGL-----GNLLHYD 50
| | | | |
QY 397 FOGQPFIMCMLHNFQGN-----HGLFGALEA-VNGPPEARL--FPNSTVW 439
| | | | |

```

```
Db 51 FONTFYCFDQLRRRFGDVSLQAMTPVVLINGLAAREALVTHGEDTADRPVPIQIL 110
QY 440 GTGMAPEGI-----SONEVVYSIMAEIGWRKDPVPDLAAMVTSFAARVGVSHD 489
Db 111 GFGPRSQGVFLARYGPAMREQRRFSVSTIRNIGLQK---KSLBQWTEBA----- 158
QY 490 AGAAMRLLIRSVYNGSGEACRGHNSPLVRPSSLQNTSIVYNSDVFEAMRLLITSAPS 549
Db 159 -----CLCAAFANHSGRPF--RPNGLDKAV---SNVI-----AS 188
QY 550 IATSPAERYD-----LIDLTRQAVQELVSYEEARSAYLSKELASILRAGVLAWE-- 601
Db 189 LTGCRREFYDDPRFLFLDLAQGLKESGFLREVINAVPVLHLPAL--AGKVLRFQKA 246
QY 602 LEPALDEVL-----ASDSRFLIGSVLBOARAAVSEAEADFEQNSRYQTLTGPE 652
Db 247 FLTQDELLETHEMTWDPAQPPRDLTEAFIAMEKAK--GNPESFNDENLR----- 296
QY 653 GNILDYANKQAGLVANYTTPRMRLEALVDSVAQGIPTQOHOQFDKXVFQLEQ----- 706
Db 297 ---IVVADLFSAGWTTSTTLAMGLMLHPDVQRRV---QOEIDVIGQVRPEMGDQ 350
QY 707 -----AFVLSKQRYPSQPRGDTVDL 726
Db 351 AHMPYTTAVIHEVQRF-----GDIYPL 372
```

Search completed: February 13, 2004, 16:19:09  
Job time : 63 secs